

THE CIPP MODEL FOR EVALUATION

- *an update*
- *a review of the model's development*
- *a checklist to guide implementation*

by

Daniel L. Stufflebeam
Harold and Beulah McKee Professor of Education and
Distinguished University Professor
Western Michigan University

Presented at the 2003 Annual Conference of the Oregon Program Evaluators Network (OPEN)

Portland, Oregon
10/03/2003

Oregon's evaluators have a long history of excellent evaluation service and creative contributions to evaluation methodology. During the 1960s, 70s, and 90s, I enjoyed and learned much from evaluation assignments in Oregon. I especially recall involvements with the Portland Public Schools, Northwest Regional Educational Laboratory, Center for Advanced Study of Educational Administration, Oregon System of Mathematics Education, and Teaching Research Division at Western Oregon University. Reflections on site visits to eastern and western Oregon, through the Willamette Valley, and along the Oregon coast evoke vivid memories of Oregon's varied and beautiful terrain. A sight I will never forget occurred near Bern when a herd of wild horses ran on either side of my moving rental car as they crossed from the mountain slope on my left to the one on my right.

I readily welcomed the invitation to participate in this 2003 conference of the Oregon Program Evaluators Network. I have chosen this venue to present an update of the CIPP Evaluation Model, an explanation of how and why it was developed, and an updated checklist for use in carrying out CIPP-guided evaluations. I hope this paper will be useful to Oregon's evaluators as they confront the varied and difficult challenges of evaluation assignments.

The CIPP Model is a work in progress. After sketching the model's current state, I will describe its origins and development, taking account of key contributing factors. These include developing and directing *The Evaluation Center*; directing or consulting on a wide range of *evaluation projects*; leading development of *professional standards for program and personnel evaluations*; conceptualizing and applying *metaevaluation*; characterizing, classifying, and assessing *evaluation models*; collaborating and deliberating with *leading evaluation theorists and practitioners* and *evaluation-oriented administrators*; studying and assisting *institutionalization of evaluation*; conducting research and development on *personnel evaluation*;

developing *evaluation checklists*; and designing and directing *evaluation masters* and *Ph.D. programs*. I will conclude the paper by describing the appended, detailed checklist for use in designing, guiding, and assessing CIPP evaluations.

PART I: THE CIPP MODEL, CIRCA 2003

The CIPP Model's current version (Stufflebeam, 2002-a, 2003-a; Stufflebeam, Gullickson, & Wingate, 2002) reflects prolonged effort and a modicum of progress to achieve the still distant goal of developing a sound evaluation theory, i.e., a *coherent set of conceptual, hypothetical, pragmatic, and ethical principles forming a general framework to guide the study and practice of evaluation*.

The CIPP Model is a comprehensive framework for guiding formative and summative evaluations of projects, programs, personnel, products, institutions, and systems. The model is configured for use in internal evaluations conducted by an organization's evaluators, self-evaluations conducted by project teams or individual service providers, and contracted or mandated external evaluations. The model has been employed throughout the U.S. and around the world in short-term and long-term investigations—both small and large. Applications have spanned various disciplines and service areas, including education, housing and community development, transportation safety, and military personnel review systems.

Context, Input, Process, and Product Evaluations

The model's core concepts are denoted by the acronym CIPP, which stands for evaluations of an entity's context, inputs, processes, and products. *Context evaluations* assess needs, problems, assets, and opportunities to help decision makers define goals and priorities and help the broader

group of users judge goals, priorities, and outcomes. *Input evaluations* assess alternative approaches, competing action plans, staffing plans, and budgets for their feasibility and potential cost-effectiveness to meet targeted needs and achieve goals. Decision makers use input evaluations in choosing among competing plans, writing funding proposals, allocating resources, assigning staff, scheduling work, and ultimately in helping others judge an effort's plans and budget. (I see input evaluation as the most neglected, yet critically important type of evaluation.) *Process evaluations* assess the implementation of plans to help staff carry out activities and later help the broad group of users judge program performance and interpret outcomes. *Product evaluations* identify and assess outcomes—intended and unintended, short term and long term—both to help a staff keep an enterprise focused on achieving important outcomes and ultimately to help the broader group of users gauge the effort's success in meeting targeted needs.

In the formative case—where evaluation helps guide an effort—context, input, process, and product evaluations respectively ask, What needs to be done? How should it be done? Is it being done? Is it succeeding? The evaluator submits interim reports addressing these questions to keep stakeholders informed about findings, help guide decision making, and strengthen staff work.

In finalizing a summative report, the evaluator refers to the store of context, input, process, and product information and obtains additionally needed information. The evaluator uses this information to address the following retrospective questions: Were important needs addressed? Was the effort guided by a defensible plan and budget? Was the service design executed competently and modified as needed? Did the effort succeed?

Partitioning the Product Evaluation Component

In summing up long-term evaluations, the product evaluation (Did it succeed?) component may be divided into assessments of impact, effectiveness, sustainability, and transportability. These product evaluation subparts ask, Were the right beneficiaries reached? Were their targeted needs met? Were the gains for beneficiaries sustained? Did the processes that produced the gains prove transportable and adaptable for effective use elsewhere?

Focus on Improvement

The CIPP Model emphasizes that *evaluation's most important purpose* is not to prove, but to improve—an idea originally put forward by Egon Guba when serving on the Phi Delta Kappa National Study Committee on Evaluation (Stufflebeam et al., 1971).¹ Evaluation is thus conceived primarily as a functional activity oriented in the long run to stimulating, aiding, and abetting efforts to strengthen and improve enterprises. However, the model also posits that some programs or other services will prove unworthy of attempts to improve them and should be terminated. By helping stop unneeded, corrupt, or hopelessly flawed efforts, evaluations serve an improvement function through assisting organizations to free resources and time for worthy enterprises.

Prospective and Retrospective Applications of the CIPP Model

Consistent with its improvement focus, the CIPP Model places priority on guiding the planning and implementation of development efforts. The model's intent is thus to supply evaluation users—such as policy boards, government officials, foundation presidents and staff members,

¹ Whereas Dr. Guba stated, “The purpose of evaluation is not to prove, but to improve,” I subsequently revised this notion, stating that “Evaluation’s **most** important purpose is not to prove, but to improve.” This change reserves a place for causal/proving functions as well as the dominant focus on improvement.

project staffs, school administrators, curriculum developers, city planners, military leaders, curriculum specialists, teachers, and counselors—with timely, valid information of use in identifying an appropriate area for development; formulating sound goals, activity plans, and budgets; successfully carrying out work plans; periodically deciding whether and, if so, how to repeat or expand an effort; and meeting a funder’s accountability requirements.

The CIPP Model also provides for conducting retrospective, summative evaluations to serve a broad range of stakeholders. Potential consumers need summative reports to help assess the quality, cost, utility, and competitiveness of products and services they might acquire and use. Other stakeholders might want evidence on what their tax dollars or other types of support yielded. If evaluators effectively conduct, document, and report formative evaluations, they will have much of the information needed to produce a defensible summative evaluation report. Such information will also prove invaluable to those outsiders engaged to conduct a summative evaluation of a given entity.

Table 1 summarizes uses of the CIPP Model for both formative and summative evaluations. The matrix’s eight cells encompass much of the evaluative information required to guide enterprises and produce summative evaluation reports. Beyond context, input, process, and product evaluations—set in both formative and summative contexts—the CIPP Model includes several other key features.

The Values Component

Figure 1 summarizes the CIPP Model’s basic elements in three concentric circles and portrays the central importance of defined values. The inner circle denotes the core values that should be identified and used to ground a given evaluation. The wheel surrounding the values is divided

Table 1. The Relevance of Four Evaluation Types to Formative and Summative Evaluation Roles

EVALUATION ROLES	<i>Context</i>	<i>Input</i>	<i>Process</i>	<i>Product</i>
<i>Formative Evaluation:</i> Prospective application of CIPP information to assist decision making and quality assurance	Guidance for identifying needed interventions and choosing goals (based on assessing needs, problems, assets, and opportunities).	Guidance for choosing a program or other strategy (based on assessing alternative strategies and resource allocation plans) followed by examination of the work plan.	Guidance for implementing the work plan (based on monitoring and judging activities and periodic evaluative feedback).	Guidance for continuing, modifying, adopting, or terminating the effort (based on assessing outcomes and side effects).
<i>Summative Evaluation:</i> Retrospective use of CIPP information to sum up the program's merit, worth, probity, and significance	Comparison of goals and priorities to assessed needs, problems, assets, and opportunities.	Comparison of the program's strategy, design, and budget to those of critical competitors and to the targeted needs of beneficiaries.	Full description of the actual process and record of costs. Comparison of the designed and actual processes and costs.	Comparison of outcomes and side effects to targeted needs and, as feasible, to results of competitive programs. Interpretation of results against the effort's assessed context, inputs, and processes.

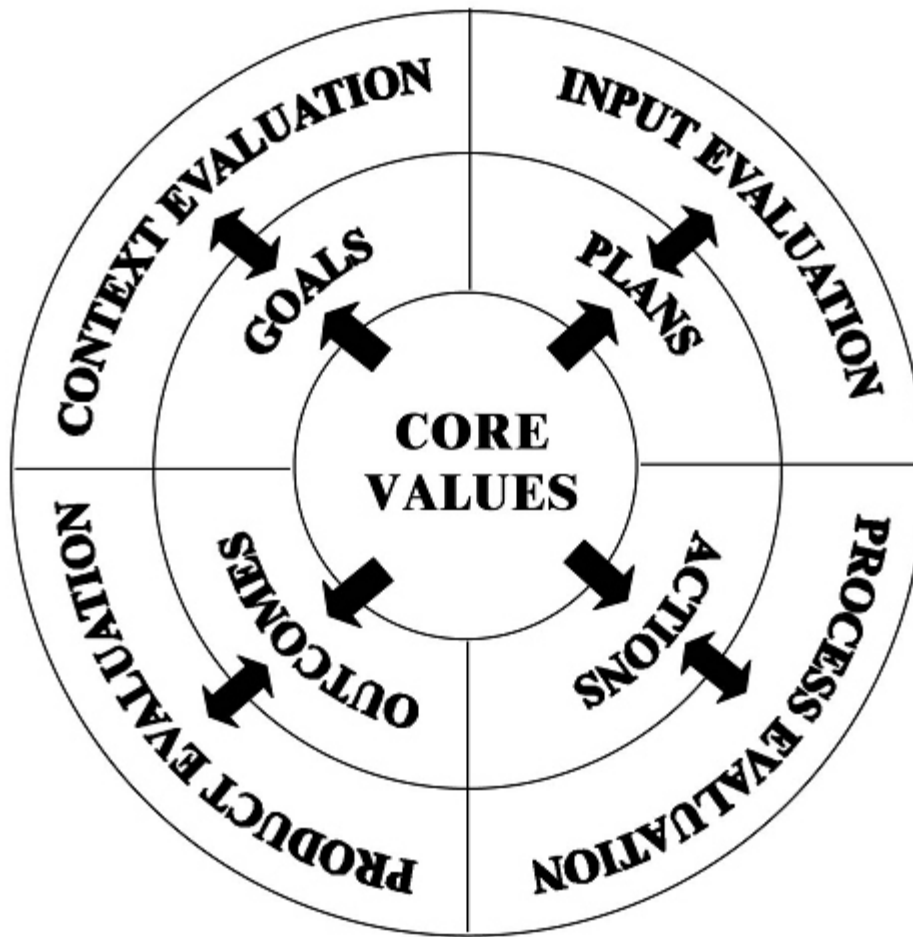


Figure 1. Key Components of the CIPP Evaluation Model and Associated Relationships with Programs

into four evaluative foci associated with any program or other endeavor: goals, plans, actions, and outcomes. The outer wheel indicates the type of evaluation that serves each of the four evaluative foci, i.e., context, input, process, and product evaluation. Each two-directional arrow represents a reciprocal relationship between a particular evaluative focus and a type of evaluation. The goal-setting task raises questions for a context evaluation, which in turn provides

information for validating or improving goals. Planning improvement efforts generates questions for an input evaluation, which correspondingly provides judgments of plans and direction for strengthening plans. Program actions bring up questions for a process evaluation, which in turn provides judgments of activities plus feedback for strengthening staff performance.

Accomplishments, lack of accomplishments, and side effects command the attention of product evaluations, which ultimately issue judgments of outcomes and identify needs for achieving better results.

These relationships are made functional by grounding evaluations in core values, referenced in the scheme's inner circle. Evaluation's root term *value* refers to any of a range of ideals held by a society, group, or individual. The CIPP Model calls for the evaluator and client to identify and clarify the values that will guide particular evaluations. Example values—applied in evaluations of U.S. public school programs—are success in helping all students meet a state's mandated academic standards, helping all children develop basic academic skills, helping each child fulfill her or his potential for educational development, assisting and reinforcing development of students' special gifts and talents, upholding human rights, meeting the needs of disabled and underprivileged children, developing students as good citizens, assuring equality of opportunity, effectively engaging parents in the healthy development of their children, attaining excellence in all aspects of schooling, conserving and using resources efficiently, assuring safety of educational products and procedures, maintaining separation of church and state, employing research and innovation to strengthen teaching and learning, and maintaining accountability. Essentially, evaluators should take into account a set of pertinent societal, institutional, program, and professional/technical values when assessing programs or other entities.

The values provide the foundation for deriving and/or validating particular evaluative

criteria. Example criterial areas pertaining to programs for students may include indicators of intellectual, psychological, aesthetic, social, physical, moral, and vocational development. Selected criteria, along with stakeholders' questions, help clarify an evaluation's information needs. These, in turn, provide the basis for selecting/constructing the evaluation instruments and procedures, accessing existing information, and defining interpretive standards.

Also, a values framework provides a frame of reference for detecting unexpected defects and strengths. For example, through broad values-oriented surveillance, an evaluator might discover that a program excels in meeting students' targeted academic needs but has serious deficiencies, such as racist practices, unsafe equipment, alienation of community members with no children in school, "burnout" of teachers, and/or graft. On the positive side, examination of a program against a backdrop of appropriate values might uncover unexpected positive outcomes, such as strengthened community support of schools, invention of better teaching practices, and/or more engaged and supportive parents.

Evaluators and their clients should regularly employ values clarification as the foundation for planning and operationalizing evaluations and as a template for identifying and judging unexpected transactions and results. Referencing appropriate values is what sound evaluation is all about. Grounding evaluations in clear, defensible values is essential to prevent evaluations from aiding and abetting morally wrong, unethical actions and instead to help assure that the evaluations will be instrumental in effectively pursuing justifiable ends. I wish to underscore that the CIPP Model is fundamentally a values-oriented model.

Evaluation Definitions

According to the CIPP Model, an evaluation is a *systematic investigation of the value of a program or other evaluand*. Consistent with this values-oriented definition, the CIPP Model

operationally defines evaluation as *a process of delineating, obtaining, reporting, and applying descriptive and judgmental information about some object's merit,² worth,³ probity,⁴ and significance⁵ in order to guide decision making, support accountability, disseminate effective practices, and increase understanding of the involved phenomena.*

Standards for Evaluations

The bases for judging CIPP evaluations are pertinent professional standards, including the Joint Committee (1988, 1994, 2003) standards for evaluations of personnel, programs, and students. These require evaluations to meet conditions of *utility* (serving the information needs of intended users), *feasibility* (keeping evaluation operations realistic, prudent, viable, and frugal), *propriety* (conducting evaluations legally, ethically, and with due regard for the welfare of participants and those affected by results), and *accuracy* (revealing and conveying technically sound information about the features that determine the evaluand's merit, worth, probity, and/or significance). See also the AEA *Guiding Principles for Evaluators* (Shadish, Newman, Scheirer, & Wye, 1995) and the U.S. General Accounting Office (2003) *Government Auditing Standards*.

The common conception that the CIPP Model is strictly focused on utilization of findings is not surprising but misleading. True, early renditions of the model attempted to replace the many ritualistic, post hoc, often experimental design-oriented, and nonconsequential educational evaluations of the late 1960s and early 1970s with evaluations that would help educators plan,

² Merit denotes something's intrinsic quality or excellence, irrespective of its utility.

³ Worth refers to something's intrinsic quality and its extrinsic value, especially its utility in meeting targeted needs.

⁴ Probity denotes something's uncompromising adherence to moral standards, such as freedom, equity, human rights, and honesty.

⁵ Significance includes but looks beyond something's intrinsic quality and utility in meeting needs to gauge the reach, importance, and visibility of the enterprise's contributions and influence.

launch, carry through, and be accountable for successful projects. Also, I continue to believe that if a contemplated evaluation has no prospect for use, it should not be initiated, whatever its potential to satisfy conditions of accuracy, propriety, and feasibility. However, for those evaluations judged worthy of implementation, the CIPP Model requires evaluators to meet conditions of accuracy, propriety, and feasibility, as well as utility. In fact, I would pan any evaluation that failed the individual propriety standard of *service orientation* (i.e., assessing whether an evaluand is protecting the interests of and addressing and effectively serving the needs of targeted beneficiaries) or any of the accuracy standards of *valid information*, *justified conclusions*, or *impartial reporting*—whether or not it met most or all of the other Joint Committee Standards.

Involvement of Stakeholders

The CIPP Model is strongly oriented to involving and serving an enterprise’s stakeholders. While evaluators must control the evaluation process to assure its integrity, CIPP evaluations accord program beneficiaries and other stakeholders more than a passive recipient’s role. Evaluators are charged to keep stakeholders informed and provide them appropriate opportunities to contribute.

Consistent with the Joint Committee’s program, personnel, and student evaluation standards (1988, 1994, 2003) and writings of Alkin, Daillak, & White (1979); Guba and Lincoln (1989); House & Howe (2000); Patton (2000); and Stake (1983), evaluators using the CIPP Model are expected to search out all relevant stakeholder groups and engage at least their representatives in communication and consensus-building processes to help affirm foundational values; define evaluation questions; clarify evaluative criteria; contribute needed information; and assess evaluation reports.

Involving all levels of stakeholders is considered ethically responsible, because it equitably empowers the disadvantaged as well as the advantaged to help define the appropriate evaluation questions and criteria, provide evaluative input, and receive and use evaluation findings. Involving all stakeholder groups is also seen as wise, because sustained, consequential involvement positions stakeholders to contribute information and valuable insights and inclines them to study, accept, value, and act upon evaluation findings.

Institutionalization of Evaluation

The CIPP Model advises evaluators to use contracted evaluations to encourage and assist evaluation clients to learn evaluation concepts and methods and install or strengthen institutional capacity to conduct and use evaluations (Guba & Stufflebeam, 1970; Stufflebeam, 1982, 1997-a, 2003-b; Stufflebeam & Webster, 1988). While external, contracted evaluations are often warranted, they are insufficient to fulfill all of an organization's ongoing requirements for evaluation, e.g., informing decision making, maintaining accountability, and fostering institutional learning. Institutions need the capacity to conduct many of their own evaluations and external evaluators should help develop such capacity.

Relevant evaluation capacity-building services—that external evaluators might deliver while conducting an external evaluation—include training institutional staff members in the concepts, standards, and methods of evaluation; engaging clients to critique draft evaluation reports; helping the institution define evaluation roles for staff; helping develop an evaluation resource library; helping organize an evaluation database; and helping draft an institutional evaluation policies and procedures handbook (See Adams, 1971; Stufflebeam, 2003-b; also see the U.S. General Accounting Office, 2003, government auditing standards chapter on attestation engagements). The payoffs from contracted evaluations are increased when evaluators and

institutional personnel use them to strengthen a client organization's commitment to and capabilities to conduct some of its needed evaluations.

A Few Words About Empowerment

Using evaluations to promote and assist institutionalization of evaluation is the sense in which I endorse evaluators' engagement in empowerment activities. However, while it is appropriate for a contracted external evaluator to help organizations build and apply evaluation capacity, the evaluation contractor must not give clients authority to determine the contents of contracted external evaluation reports or in any way corrupt evaluation processes and reports (See Stufflebeam, 1997-b, 1999-a, 1999-b, 2000). In training clients to conduct evaluations, evaluation contractors should stress that an organization's self-evaluations should evidence utmost integrity by following and giving evidence of meeting professional standards of sound evaluation, including especially controlling conflicts of interest and bias.

Objectivist Orientation

The CIPP Model's epistemological orientation is objectivist rather than relativist. Objectivist evaluations are based on the theory that moral good is objective and independent of personal or human feelings. Such evaluations are firmly grounded in ethical principles, such as the U.S. Bill of Rights; strive to control bias, prejudice, and conflicts of interest in conducting assessments and reaching conclusions; invoke and justify appropriate and (where they exist) standards of technical merit; obtain and validate findings from multiple sources; search for best answers, although these may be difficult to find; set forth and justify best available conclusions about the evaluand; report findings honestly, fairly, and as circumspectly as necessary to all right-to-know audiences; subject the evaluation process and findings to independent assessments against pertinent standards; and identify needs for further investigation. Fundamentally, objectivist

evaluations are intended, over time, to lead to conclusions that are correct—not correct or incorrect relative to an evaluator’s or other party’s predilections, position, preferences, standing, or point of view. The CIPP Model contends that when different objectivist evaluations are focused on the same object in a given setting, when they are keyed to fundamental principles of a free and just society and agreed-upon criteria, when they meaningfully engage all stakeholder groups in the quest for answers, and when they conform to the evaluation field’s standards, different, competent evaluators will arrive at fundamentally equivalent, defensible conclusions.

Use of Multiple Methods

The CIPP Model requires engagement of multiple perspectives, use of a wide range of qualitative and quantitative methods, and triangulation procedures to assess and interpret a multiplicity of information. Given the emergent nature and dynamic environments of many evaluands, assessed, common laboratory controls usually are not feasible; pertinent, validated data gathering instruments often do not exist; and typically there is too little time to thoroughly design, pilot test, and validate the needed instruments. Consequently, the evaluator has to be resourceful in compiling a wide range of reasonably good information that in the aggregate tells a consistent, truthful story. The model advocates engaging multiple observers and informants with different perspectives; constructing “homemade” instruments as needed; mining and using extant, pertinent information; addressing each evaluation question in a timely manner; using multiple procedures; cross-checking qualitative and quantitative findings; building a compelling case over time; and subjecting the evaluation to review by stakeholder groups and independent parties. In following this advice, evaluators are expected to search out and investigate ambiguities and convergence and contradictions in findings, listen to and weigh feedback from

the program's stakeholders, and be appropriately circumspect in generating and reporting conclusions.

Table 2 illustrates the variety of methods of use in evaluations of context, input, process, and product (with product evaluation divided into the subparts of impact, effectiveness, sustainability, and transportability). Almost all the methods listed down the table's vertical axis apply to more than one type of evaluation. As indicated by multiple checkmarks in each column, use of multiple methods for each type of evaluation provides needed crosschecks on findings. Implementing a many-faceted approach to methodology often requires that a team conduct an evaluation. Collectively the team should possess such competencies as knowledge of the pertinent subjective matter, planning, negotiation and contracting, leading groups, organizing and administering team efforts, using technology, interviewing, surveying, testing, quantitative and qualitative analysis, cost analysis, effective writing, and effective oral communication.

Contracting

Early applications of the CIPP Model revealed the critical importance of contracting (for example, see House, Rivers, and Stufflebeam, 1974). All evaluations have the potential for the emergence of misunderstandings and disputes concerning a wide range of matters. These can concern authority to edit reports and release findings to various potential audiences, whether or not there will be a prerelease review of reports, what deliverables are agreed to, what amounts of money will be paid upon completion of various tasks, when evaluation reports are due, etc. Evaluators and their clients should discuss and reach agreements on such matters before proceeding with an evaluation. A signed contract provides a basis for settling later possible disputes or renegotiating previous agreements.

Table 2. Illustration of Methods of Potential Use in CIPP Evaluations

METHODS	<i>Context</i>	<i>Input</i>	<i>Process</i>	<i>Impact</i>	<i>Effectiveness</i>	<i>Sustainability</i>	<i>Transportability</i>
Survey	X		X	X	X	X	
Literature Review	X	X					
Document Review	X	X	X	X	X		
Visits to Other Programs		X		X	X		X
Advocate Teams (to create & assess competing action plans)		X					
Delphi Technique	X	X					
Program Profile/Database		X	X	X	X	X	
On-Site Observer			X	X	X	X	
Case Studies			X	X	X	X	
Comparative/ Experimental Design Studies		X			X	X	
Stakeholder Interviews	X		X	X	X	X	X
Focus Groups	X	X	X	X	X	X	X
Hearings	X	X			X		
Cost Analysis		X	X		X	X	
Secondary Data Analysis	X				X		
Goal-Free Evaluation			X	X	X	X	X
Photographic Record	X		X	X	X	X	X
Task Reports/Feedback Meetings	X	X	X	X	X	X	X
Synthesis/Final Report	X	X	X	X	X	X	X

Guidance for setting up such contracts is available in checklists by Robert Stake (2001) and Stufflebeam (2001-a) at <www.wmich.edu/evalctr/checklists>. (Also, see Stufflebeam, 1999-a and 2000.) While a formal contract may seem out of place in relatively informal, internal types of evaluation, it is still good practice at least to work out an up-front, letter of agreement.

Design Elements

Designing a CIPP evaluation initially involves setting up a comprehensive plan keyed to stakeholder questions and pertinent criteria of merit, worth, probity, and/or significance. While initial design decisions are needed to fund, staff, and launch an evaluation, they should not be considered fixed. Instead the CIPP Model treats evaluation design as a process of continually identifying and employing appropriate means to address emergent as well as predictable and relatively fixed information needs.

I have found that four particular matrixes are useful in compiling and communicating CIPP evaluation designs. The first matrix arrays the evaluation methods to be employed in a given evaluation along the vertical axis and the context, input, process, impact, effectiveness, sustainability, and transportability components of the CIPP Model along the horizontal axis. Checkmarks in pertinent cells show which methods will address each CIPP component and collectively indicate the extent of methodological checks and balances for each CIPP component.

A second matrix also arrays the selected evaluation methods along the vertical axis but shows each evaluation year (or month) along the horizontal axis. Checkmarks in the appropriate cells show which evaluation methods will be employed in each month or year. This type of analysis is invaluable for communicating and negotiating the schedule of information needs and for managing data collection. The third matrix arrays the selected evaluation methods down the vertical axis and planned evaluation reports across the horizontal axis. Checkmarks in the appropriate cells show which methods will contribute information for which reports.

The fourth matrix arrays planned evaluation reports down the vertical axis and evaluation years (or months) across the horizontal axis. Checkmarks in the appropriate cells indicate which reports will be presented in which years or months.

Deciding Which Parts of the CIPP Model to Employ and What New Information to Collect

Must one always carry out all four types of evaluation, i.e., context, input, process, and product? is a frequent question about the CIPP Model. The answer is no. Deciding which parts of the model to employ is a critically important task in evaluation planning and depends on three main factors.

The first concerns when in the life of a program the evaluation is introduced. In the rare case when the evaluator is called in before the program begins, potentially the evaluators would conduct all four kinds of evaluation to help guide the program through focusing, planning, implementation, and success at the end. In the more typical, but often unfortunate case where the evaluator is called in only after the program has been completed, one might also sum up the program's value by looking at context, input, process, and product. Sometimes one is engaged to evaluate somewhere in the program's middle. Then one might, at least for the time being, hold context and input evaluation in abeyance and concentrate on process and product evaluation.

However, in any determination of which types of evaluation to apply, the evaluator needs to identify and address the client's purpose for the evaluation. A summative evaluation will almost always require all four types of evaluation in order to fully describe the program and appropriately judge its quality. However, a formative evaluation assignment sometimes will take up only the type(s) of evaluation needed to guide certain program decisions or answer pointed accountability questions.

In one evaluation the U.S. Marine Corps requested only an input evaluation, because the Corps had already assessed and decided to replace its existing system for evaluating officers and enlisted personnel. The purpose of the input evaluation was to critically compare personnel review systems used by other military organizations and three new proposed systems. In this case, we didn't need to collect a lot of context evaluation information—related to needs and problems in the existing personnel review system—because the Corps had already done so. We worked from the Corps' context evaluation information and added what we saw as missing. Also, following the choice of a plan for a new system, the Corps conducted its own process and product evaluation in the course of launching the new system. In delivering our final reports, we provided process and product evaluation designs that the Corps could employ in the course of installing its new system.

As seen in the Marine Corps case, a third factor to be considered when deciding what kind of evaluation to conduct concerns how much of the information needed for context, input, process, and product evaluation is already available or is potentially available from other evaluations of the program. The point is that one's program evaluation is not always the "only game in town." In assessing context, input, process, and product, one should compile the information required by each pertinent type of evaluation. One should not redundantly gather new information if acceptable and sufficient information from another source is readily available. On this point, when starting an evaluation I always acquire and study the extant, pertinent information as an aid to deciding what new information to gather. Such information may appear in various forms, e.g., past evaluation reports, needs assessment sections of funding proposals, doctoral dissertations, accountability reports to sponsors, newspaper accounts, etc.

Interim Reports

Throughout most CIPP evaluations the evaluator should have regular exchanges with the client group to present and discuss recent findings and, as appropriate, review and modify the evaluation design and schedule.

The Feedback Workshop Technique (Gullickson & Stufflebeam, 2001) has proved valuable for delivering and discussing formative evaluation findings and updating evaluation designs. This technique involves the evaluator in periodically convening with the client and selected stakeholders. About two weeks before each meeting, the evaluator sends all members of the client group drafts of the most recent reports along with a request that they identify any factual errors or instances of ambiguity. The recipients read and critique the reports in preparation for the meeting. At the meeting the evaluator briefs the client group on the evaluation findings, using handouts, transparencies, a computerized presentation, etc.. The client group members then respond to the report and raise their questions, followed by discussion with the evaluator. The evaluator subsequently notes any corrections to be made to the evaluation reports, projects the evaluation's next steps, invites reactions and suggestions regarding planned evaluation activities, and requests any needed assistance regarding future data collection activities. The evaluator also presents the client with audiovisual materials that he or she might use to brief constituents on the most recent evaluation findings. It is useful to conclude the session by briefly engaging each participant to state the most important points or issues in the preceding exchange.

Following the feedback workshop the evaluator or another workshop participant prepares and distributes a summary of the meeting. The evaluator also finalizes the discussed reports,

updates the evaluation design, and sends copies of both to the client, who will distribute them further as appropriate.

Summing Up and Reporting Evaluation Findings

At an enterprise's attainment of maturity or completion, the CIPP Model sees the evaluator compiling all relevant information in a comprehensive summative evaluation report. Consistent with the model's objectivist orientation, the evaluator should seek convergence on conclusions about the evaluand's value and lessons learned. To address questions of the full range of audiences, the summative report should assess the evaluand's context, inputs, process, and products and attend to stakeholders' high priority questions. Moreover, the evaluator should be resourceful in effectively and efficiently reporting to the different audiences.

An illustrative outline for a summative report designed to meet the differential and common needs of a disparate audience follows:

Part 1: Program Background (The Operating Organization, Genesis of the Program, the Program's Environment)

Part 2: Program Implementation (Program Overview, Program Beneficiaries, Program Financing, Program Governance, Program Staff, Program Facilities, Program Operations)

Part 3: Program Results (Evaluation Design, Findings [context, inputs, process, impact, effectiveness, sustainability, transportability], Conclusions [judgments of merit, worth, probity, and significance, and lessons learned]).

In this example, Part 1 is descriptive and targeted to persons lacking information on the operating organization, why it started the program, and the program's locale. The mainly descriptive Part 2 is mainly descriptive and directed to those who might be interested in

replicating part or all of the program. Part 3 presents a comprehensive appraisal of the enterprise and its outcomes and is targeted to all members of the audience.

Beyond these three parts, the summative evaluation report should have a sufficient appendix. It should help readers learn and judge who conducted the evaluation, how independent and qualified they were, how they carried out all important aspects of the evaluation, succinctly what they found, and how the evaluation work and products were judged. Example components for the appendix to a summative evaluation report include a list of interim reports, field evaluator's handbook, sampling strategies, interview protocols, survey forms, focus group session agendas and participants, documents reviewed, goal-free evaluation plan, evaluation personnel and their qualifications, the evaluation contract or memorandum of agreement, data tables metaevaluation agents and findings, evaluator's assessment of the evaluation against pertinent standards, client's response to the evaluation, and executive summary.

Metaevaluation

The CIPP Model stipulates that evaluations themselves should be rigorously evaluated, a process referred to as metaevaluation (Scriven, 1972; Stufflebeam, 1974, 2001-b). Both formative and summative metaevaluations are needed. At a minimum the evaluators themselves should employ formative metaevaluation to guide the work and, at the end, conduct and report a summative metaevaluation. Additionally, they should encourage the evaluation's client to obtain a metaevaluation from a qualified third-party evaluator.

Evaluators need formative metaevaluations to help assure that their evaluations maintain useful focus, are soundly executed, and are appropriately and effectively serving clients. Anyone who has conducted much field-based evaluation knows that evaluation needs evolve throughout the course of a study, as new audiences are identified, as a program unfolds, as clients'

information needs change, etc. Also, many things can and do go wrong in evaluations.

Evaluators should regularly subject their ongoing evaluations to review and revise/correct evaluation procedures and actions as appropriate.

When an evaluation has been completed, audiences who find the evaluation findings and conclusions disquieting, dubious, or not totally convincing can benefit from seeing an independent assessment of the evaluation. By regularly subjecting evaluations to independent metaevaluations, evaluators can better appraise the soundness of their evaluation services, make needed improvements in future evaluations, and enhance their credibility to present and future clients.

The bottom line argument for metaevaluation concerns reciprocity. Evaluators, who expect others to have their work evaluated, likewise should subject their evaluation plans, process, and reports to creditable metaevaluations. The CIPP Model advocates use of the Joint Committee (1988, 1994, 2003) Standards and other appropriate criterial sets, especially the American Evaluation Association's *Guiding Principles* [Shadish, Newman, Scheirer, & Wye (Eds.), 1995], and the U.S. General Accounting Office (2003) *Government Auditing Standards* as the basis for metaevaluations. In educational evaluations involving testing and other assessments of students, evaluators often should apply the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

PART II: ROOTS OF THE CIPP MODEL

In this part, I turn the calendar back and consider why and how the CIPP Model was developed and acquired its current configuration. I have decided to share this information because many students of the evaluation field have increasingly been asking me and other evaluation writers to help them gain insight into the conditions that led to the different evaluation models. I agree with many who have observed that history is important in understanding, assessing, and improving on current practices.

Work on the CIPP Model began in 1965 because U.S. public schools were finding they could not meaningfully and successfully evaluate their federally supported projects using the then “gold standard” for program evaluations—controlled, variable manipulating, comparative experiments (Guba, 1966; Stufflebeam, 1966-b, 1967-a).⁶ Since its humble beginnings, the CIPP Model has been extensively developed and widely applied (For example, see Candoli, Cullen, & Stufflebeam 1997; Gally, 1984; Granger, Grierson, Quirino, & Romano, 1965; Guba & Stufflebeam, 1968; Nevo, 1974; Stufflebeam, 1969, 1995, 1997-a, 2003-b; Stufflebeam et al., 1971; Stufflebeam, Candoli, and Nicholls, 1995; Stufflebeam, Gullickson, & Wingate, 2002; Stufflebeam & Millman, 1995; Stufflebeam & Nevo, 1976; Stufflebeam, & Webster, 1988; Webster, 1975). Those employing or contracting others to employ the model have included government officials, foundation officers, program and project staffs, international assistance personnel, agricultural extension agents, school administrators, church officials, physicians, military leaders, and evaluators.

⁶ See Smith, M. F. (1989) for a delineation of the often unattainable preconditions required for proceeding with a sound field experiment.

My present view of evaluation was derived, not from conceptualizing in university ivory towers, but from working in evaluation's "school of hard knocks." My beliefs about evaluation are based on learning by doing and an ongoing effort to identify and correct mistakes made in evaluation practice. If I could offer but one recommendation to evaluation beginners, it is that they should learn evaluation throughout their careers by conducting and reflecting on a wide range of evaluations. Now let's look at some of the lessons I learned in that manner.

My Quite Accidental Entry Into Evaluation

In 1965, when I was directing The Ohio State University (OSU) Test Development Center, Dr. John Ramseyer—Head of OSU's School of Education—entered my office and said, "Dan, you're going into the evaluation business." He noted the federal government had offered huge school improvement grants to U.S. school districts under the Elementary and Secondary Education Act of 1965 (ESEA). Ohio's schools needed funds to improve the education of disadvantaged students, but couldn't meet the ESEA evaluation requirements; and—as Ohio's flagship university—OSU had to help the schools. He said Dr. Egon Guba would oversee my efforts. As a 28-year old, malleable assistant professor, I said "Yes, sir" and set off to become an "evaluator."

Although I had never taken an evaluation course, I was in a good place to start an evaluation career. Ralph Tyler—widely acknowledged as the father of educational evaluation—had created an evaluation reputation for OSU, especially through directing the famous Eight Year Study of progressive schools (Smith & Tyler, 1942). Maybe his halo and also his advice would help me, since he still came to OSU occasionally to visit his brother Keith and colleague Edgar Dale. Moreover, Dr. Ramseyer made a \$10,000 travel fund available to me so that I could visit and study sites of suspected evaluation activity.

I had some knowledge, skills, and associations of use in evaluation work. I had learned statistics and experimental design at Purdue from renowned statistician Benjamin Winer. In every undergraduate and graduate educational psychology and measurement course I'd taken at Iowa, Loyola, and Purdue, I'd read and reread the then top measurement texts by Lindquist, Cronbach, Anastasia, Guildford, and Thorndike and Hagen. Beyond quantitative assessment and analysis, I had taken nearly all of Purdue's graduate courses in clinical psychology/assessment and qualitative analysis and had experience supervising and evaluating students' performance in counseling practicums.⁷ Egon Guba, my OSU (and still) mentor, was and is a creative genius with strong research methodology skills; he and other Ohio State colleagues, including David Clark and Sydney Pressey, provided invaluable examples of how to conceptualize and analyze inquiry problems. In addition, I was developing many standardized achievement tests, researching matrix sampling (Cook & Stufflebeam, 1967; Owens & Stufflebeam, 1964) and the Program Evaluation and Review Technique (PERT) (Stufflebeam, 1967-b); and I was leading a study of innovations in Ohio schools (Heck, Stufflebeam, & Hock, 1966). Moreover, I had been given leave to attend the 1965 eight-week University of Wisconsin summer institute on experimental design and statistics—to be presented by Julian Stanley, Donald Campbell, Dick Schutz, and Gene Glass—where I would develop collegial relationships with outstanding fellow participants, including especially Jason Millman.

How could I fail in my new evaluation assignment? I could dust off Ralph Tyler's (1942) *General Statement on Evaluation*, use it to help schools write behavioral objectives, construct

⁷ Because of my extensive training in both qualitative and quantitative methods at Purdue, and because both approaches have limited capacity to address complex evaluative questions, I have never understood the argument that one should apply either quantitative methods or qualitative methods. As Julian Stanley once told me, an inquirer must "do his 'damnedest' with his mind," which I believe leads inescapably to employing both qualitative and quantitative methods.

achievement tests keyed to the objectives, follow the Campbell and Stanley (1963) advice in assisting schools to design and carry out randomized experiments, administer the achievement tests following project cycles, use what I had learned from Ben Winer to conduct analyses of variance and appropriate a posteriori tests for identifying and investigating statistically significant project outcomes, and report methodologically defensible findings. I could also consult as needed with Ohio State colleagues possessed of high-level conceptual and technical skills.

Although from 1958-1961 I had taught school in rural Iowa and Chicago, I had since become engrossed in “gullible’s travels” at graduate school and other venues. Obviously, I had forgotten that dynamic school settings typically are not amenable to laboratory research controls and that disadvantaged students have needs and problems that vary widely in kind and amount and are not well measured by standardized tests. My initial thinking about evaluating the schools’ War on Poverty projects would prove to make as little sense as does the now fashionable advocacy of “theory-based evaluation”—wherein one assumes that the complex of variables and interactions involved in running a project in the complicated, sometimes chaotic conditions of the real world can be worked out and used a priori to determine the pertinent evaluation questions and variables (Stufflebeam, 2001-b). Braybrooke & Lindblom (1963) discredited this notion 40 years ago, and their message clearly needs to be revisited.

Start-Up of The Evaluation Center

Despite my initial unrealistic approach to evaluating school projects, I took some steps in my new assignment that over time worked out pretty well. Dr. Ramseyer approved my recommendation to start a new evaluation center. The Evaluation Center’s mission became then—and nearly 40 years later is—to *advance the theory and practice of evaluation*. Since no

center could directly serve the evaluation needs of all public schools in Ohio, I projected we would conduct a few representative evaluation projects; use these as research and training laboratories; produce models, methods, and tools that schools could adopt and adapt; graduate some well-trained evaluation masters and doctoral students; and help set up a few exemplary school-district-based evaluation offices. The new Ohio State center would and (now at Western Michigan University) continues to conduct research, development, dissemination, training, and leadership in the context of its evaluation service projects. (See Stufflebeam, 2002-b for a checklist that reflects what I think I have learned about developing and directing successful university-based R & D centers.)

Evaluation of Elementary and Secondary Education Act (ESEA) Projects

The Evaluation Center's first project was for the Columbus, Ohio, school district. In 1965 the 110,000 student district potentially was entitled to receive \$6.6 million for 3-year, ESEA, Title I projects (targeted to economically disadvantaged students). However, according to Superintendent Harold Eibling, the district had no staff member who could write acceptable evaluation plans. I projected that The Evaluation Center could help the district meet the evaluation requirements, but said we would do so only under certain conditions keyed to The Evaluation Center's mission. These included (1) writing the needed evaluation plans and supervising their execution; (2) staffing the evaluations with school district personnel; (3) providing the school district's evaluators with on-the-job training and graduate education; (4) an intent by the district to fund, staff, and install its own office of evaluation at the end of three years; and (5) studying and documenting the experience. Superintendent Eibling agreed to the basic approach, but objected to conditions 2, 3, and 4. I noted that the new Center's aim was to try to work ourselves out of jobs and help create institutional models of school-district-based

evaluation that other districts could study and possibly emulate. Otherwise, I was convinced the Center's staff would fail to advance evaluation theory and practice and make little impact on districts' long-term evaluation capacities. Dr. Eibling reluctantly agreed to my conditions; and off I went on a well intentioned, but partly misguided effort.

Consultants had helped Columbus educators complete essentially fill-in-the-blanks Title I proposals, and I added the evaluation plans. Following basically "rubber stamp" funding by the government, members of my evaluation team engaged "focus groups" of project staff in a time-consuming process to clarify each project's objectives. Subsequently, we intended to develop or select the needed achievement tests and other instruments. We gave up early on randomly assigning disadvantaged students to the Title I projects and control groups, because this clearly was not feasible and likely was illegal. Also, we soon found that existing achievement tests were poor matches to the developmental needs of the targeted Columbus students and that project evaluators could not wait the two or more years required to design, construct, pilot test, revise, norm, and validate new achievement tests.

Out of the Armchair and Into the Schools

When I became bored watching the "parlor game" of project leaders writing behavioral objectives, I decided to visit schools to find out what was going on in the funded projects. Confusion was evident wherever I went. Nothing like the proposals' intended activities was happening. Those responsible for carrying out the projects had not helped write the proposals, many had not seen the plan they were supposed to be implementing, promised project resources had not been delivered, and the needs and problems in classrooms often seemed critically important but unrelated to procedures in the proposals and the behavioral objectives being

worked out at the central office. Teachers and principals often were upset with the situation and for good reason.

Rejection of the 1960s Evaluation Orthodoxy

A light bulb went on. Here was an opportunity to improve evaluation theory, if only to discredit the prevalent views. Many evaluation plans that appeared in proposals were true to the then current evaluation orthodoxy, i.e., evaluations should determine whether valued objectives had been achieved and met requirements of experimental design and post hoc, objective measurement. This conceptualization was wrong for the situations I found in Columbus classrooms. At best, following this approach could only confirm schools' failures to achieve (dubious) objectives. Such evaluations would not help schools get projects on track and successfully meet the education needs of poor kids.

After I decided to criticize current ideas about evaluation and call for a reconceptualization of evaluation, a relevant opportunity emerged. In January 1966, the Michigan Department of Education engaged me to give the keynote address at a statewide conference on evaluation of projects under Title I of ESEA. At the conference (Stufflebeam, 1966-a), I said I'd learned just enough in my evaluation of such projects that I had to reject basically everything I had thought necessary for evaluating educational projects, including behavioral objectives, experimental designs, and standardized tests.

Orienting Evaluation to Decision Making

Instead, I advised educators to key evaluations to provide information for decision making. I identified key types of decisions as those day-to-day choices involved in making projects work and the annual decisions about whether to retain, expand, or discontinue a project. For these

implementation and recycling decisions I suggested that schools concentrate, respectively, on conducting and reporting process and product evaluations.

Overall, the Michigan educators disliked what I said. As staunch supporters of local control of schools, most seemed averse to federally mandated evaluations and (possibly as supporters of the Michigan Wolverines and determined enemies of the Buckeyes from the state to the south) not receptive to advice from an Ohio State professor. However, three influential people strongly supported my analysis and asked me to assist their evaluation efforts. They were Drs. Stuart Rankin and Robert Lankton, the heads of evaluation and testing in the Detroit Public Schools, and a representative of the U.S. Office of Education.

An Opportunity to Develop and Test a New Approach to Evaluation

A few days after returning to Columbus, Dr. Ramseyer called me to his office. He related that a U.S. Office of Education official had requested that OSU release me so I could devote full time in Washington to leading the federal evaluation of Titles I and III of ESEA (a dubious decision, since in Michigan I mainly had said that everything I knew about evaluation wouldn't work in the ESEA program). Dr. Ramseyer said he had rejected the government's request but, instead, committed me to work on the assignment Mondays and Tuesdays in Washington. For the rest of 1966 and nearly all of 1967 I served two days a week in D.C.⁸ I stayed in a suite at the Congressional Hotel for \$12 a night; ate outstanding, low-cost breakfasts in the Rayburn Building cafeteria; chaired the ESEA evaluation committee; and frequently ate dinners at the Astor Restaurant where Congressional aides sipped 29 cent martinis while concocting plans for President Johnson's promised Great Society. I spent the subsequent three days of each week in

⁸ I reported to Dr. Nolan Estes, Associate Commissioner of Elementary and Secondary Education. After he left government service in 1970, I helped him and Dr. William Webster establish and develop the still renowned Research, Planning, and Evaluation Division of the Dallas Independent School District.

Columbus directing The Evaluation Center, conducting local and state-level evaluations, and teaching. (My wife, who solely cared for our three little ones during my weekly absences, still resents that part of my career; she did a great job and the kids turned out well!)

Early Reflections on Lessons Being Learned

Through the related but different assignments I gained an appreciation for the problem of keying evaluations to the very different information requirements of local, state, and national audiences. For example, I saw firsthand the stress that resulted in a heart attack to the Washington official charged with producing a national evaluation report on the first year of ESEA Title I (U.S. Office of Education, 1966). Largely to appease school personnel throughout the country, the U.S. Office of Education had allowed each school district to submit an annual evaluation report based on its own evaluation questions, methods, and instruments. The 14,000+ school districts subsequently flooded the U.S. mail system with tens of thousands of idiosyncratic evaluation reports not amenable to storage, retrieval, and reading, let alone data aggregation and summarization. The official charged with developing the report for Congress futilely attacked the impossible task of pulling the information together to answer pointed questions of Congress until he suffered a stress-induced heart attack. A lesson? An old one from information science—investigators must specify in advance questions and data requirements at each administrative level at which an audience requires answers.

Soon after beginning the weekly trips to D.C., some OSU senior professors inveigled me to conduct a universitywide symposium on the issues in evaluating ESEA projects. (These sage overseers of OSU's scholarly integrity said soberly that I needed to account for my trips to Washington so the university community could gauge whether my time was being spent productively.) Attendees at the symposium readily agreed with my contentions that laboratory

research methods wouldn't work well in school-based development projects; that evaluations should be functional, as in assisting decision making; and that evaluators should assess process as well as product. However, three professors from the philosophy department asserted that process and product evaluations are not enough. They stressed that I had failed to address the key need to evaluate project goals.

As I worried this issue later, I decided the philosophy professors were correct. Projects could go very wrong if guided by unclear or bad goals, and I had seen that happening in several ESEA projects. In keeping with my new view that evaluations should inform decisions, I decided that goal setting was critically important and should be guided by context evaluations, i.e., assessments of needs, problems, assets, and opportunities.

The CIPP Model Takes Shape

At this point, the basic structure of the CIPP Model was nearly complete. It included context evaluation to guide goal setting, process evaluation to guide project implementation, and product evaluation to guide recycling decisions. I subsequently added input evaluation to aid in planning projects, e.g., proposal writing (Stufflebeam, 1967-a).

An Invitation to Test the CIPP Model

After I had presented the model at a national evaluation conference in Florida (Stufflebeam, 1968), Edwin Hindsman, then director of the Southwest Educational Laboratory in Austin, Texas, invited me to test the model on one of the lab's major projects. His lab had been assigned to mount and evaluate a \$10 million program for meeting the educational needs of migrant children. It was agreed that Egon Guba, Robert Hammond, and I would use the CIPP Model to help the lab evaluate the migrant education program. Among the lessons learned from this rich experience were that the CIPP Model has to be applied flexibly [e.g., what Bob Stake

(1983) terms “responsively”]; active members of the migrant community provided more cogent information on educational and related needs of migrant kids than did experts who had been studying migrant children; and influence of decisions requires much more than submission of periodic written reports.

We also gained insights into input evaluation. To guide the needed input evaluation, we invented the *advocate teams technique* (see Reinhard, 1972) through which competing teams develop proposals for meeting a set of targeted needs (e.g., strongest possible program for addressing the needs of migrant students). Evaluators then assess the alternatives’ merits on predetermined criteria, and a convergence team may subsequently merge the best features of competing proposals into a single plan.

The PDK Book

In 1969 Phi Delta Kappa International (PDK) engaged me to head a national study committee on evaluation, which culminated in the book *Educational Evaluation and Decision Making* (Stufflebeam et al., 1971). That book sharply criticized the traditional views of educational evaluation, analyzed the evaluative information needs in decision making, and elaborated the CIPP Model. The book also suggested that criteria for judging evaluations should include utility and feasibility as well as technical adequacy. We noted that evaluations can go very wrong if keyed exclusively to criteria of technical adequacy, such as the requirements for internal and external validity then being promulgated for judging experiments (Campbell & Stanley, 1963). (The PDK book’s breakout of utility criteria into relevance, importance, timeliness, clarity, and credibility was a precursor of the work done by the Joint Committee on Standards for Educational Evaluation [Joint Committee, 1981].)

The Model Training Program in Evaluation

The PDK book garnered much interest and support and was undoubtedly a factor in The Evaluation Center winning a major competitive federal grant in excess of \$1 million in 1971 to establish and operate a model graduate training program in evaluation. A consortium of universities; school districts; state education departments; and educational research, development, and dissemination organizations operated the program. An underlying goal was to help some of the participating organizations staff and institutionalize systematic evaluation.

In its first two years the program was highly successful. It contributed to strengthening evaluation operations in such organizations as the public schools in Dallas, Texas; Detroit, Lansing, and Saginaw, Michigan; and Cincinnati, Columbus, and Xenia, Ohio. It also helped institutionalize evaluation in the Ohio State University Center for Vocational and Technical Education; regional educational laboratories in Texas and Oregon; and the state education departments in Michigan and Ohio. Among this project's products and those of early Evaluation Center training were dozens of graduates who would successfully apply evaluation in their careers and help shape the evaluation profession.

Debacle at O.S.U.

Simultaneously, however, storm clouds appeared on the home front. The Evaluation Center had grown into a relatively large, nationally visible organization, with substantial external funding. Its organizational location at Ohio State within a small educational development department proved debilitating to the Center's progress and its national leadership activities. After university officials promised but didn't follow through to locate the center in a viable location, I resigned. A lesson I took with me is that a university-based center should report to a dean or vice president. Key reasons are to avoid debilitating politics and natural conflicts of interest at the

departmental level and instead to provide a home where a center can do what it needs to do, control its grant funds, and draw participation from across the university's array of disciplines.

Taking the Work to Michigan

After assessing job invitations from various universities and research institutes across the country, in 1973 I took The Evaluation Center's mission, another staff member, four of its students, and a small amount of grant funds to Western Michigan University (WMU). A key consideration in this choice was that the Center would report to WMU's Dean of Education, Dr. John Sandberg. I had developed a high regard for Dr. Sandberg's leadership qualities when I had evaluated his previous work as Deputy Director of the Northwest Regional Education Laboratory. Another deciding factor in my move to Michigan was that WMU was located where my kids could continue their figure skating and ice hockey activities.

The Center's work prospered at WMU, and a number of Center involvements influenced my further development of the CIPP Model. Because this paper is already long, I will list and summarize only a few of the major experiences at WMU that contributed to my current views of evaluation.

1. Through a strong, sustained collegial relationship with Michael Scriven plus assignments to conduct several external evaluations, I strengthened the CIPP Model's attention to summative evaluation and affirmed and explained my objectivist philosophy of evaluation.
2. As founding chair of the Joint Committee on Standards for Educational Evaluation, I learned firsthand the virtues and difficulties of engaging a diverse group of evaluators and evaluation users to collaborate in defining what is meant by sound evaluation. While I value diverse perspectives in the course of an evaluation, I also believe that after

conducting what Cronbach et al. (1981) referred to as evaluation's divergent stage, the evaluator should skillfully lead evaluation participants to converge on a conclusion about an evaluand's merit, worth, probity, and significance.

3. I have been fortunate over the years to interact and work with leading evaluation theorists and methodologists, including especially Egon Guba, Richard Jaeger, Thomas Kellaghan, George Madaus, Jason Millman, Michael Scriven, Robert Stake, Ralph Tyler, and William Webster. These relationships helped me clarify where my views of evaluation agree and differ with other views (e.g., see Stufflebeam, Madaus, & Kellaghan, 2000; Stufflebeam, 2001-c). Also, I came to understand that differences in evaluation models often are a function of the theorists' different evaluation experiences and contexts, as well as their philosophical differences. For example, my initial and continuing commitment to make evaluations assist improvement stems from my early involvement with innovative, developing/floundering projects in schools. On the other hand, Michael Scriven's early and continuing commitment to help consumers judge completed products, no doubt stems from his early involvement in evaluating completed national curriculum packages that had been put out for public use and were past the need for formative evaluation. In general, I have made a continuing study of alternative evaluation models to compare their strengths and weaknesses and areas of applicability (Stufflebeam, 2001-c)
4. Through contentious metaevaluations conducted on the Michigan Educational Assessment Program (House, Rivers, & Stufflebeam, 1974) and on the National Assessment of Educational Progress (Stufflebeam, Jaeger, & Scriven, 1992), I asserted and demonstrated the importance of negotiating a clear, sound contract before proceeding with an evaluation (Stufflebeam, 2000).

5. As the designer and director of the National Center for Research on Educational Accountability and Teacher Evaluation (CREATE), I extended the CIPP Model for use in evaluating teachers, superintendents, and other school personnel (Shinkfield & Stufflebeam, 1995, Candoli, Cullen, & Stufflebeam, 1997).
6. In directing longitudinal evaluations of welfare housing and community and economic development in the Philippines, Hawaii, and Chicago, I extended applicability of the CIPP Model to fields outside education and explicated product evaluation into subparts of impact, effectiveness, sustainability, and transportability.
7. In leading an evaluation of the Marine Corps system for evaluating officers and enlisted personnel, I gained greater insights into roles that clients can play as evaluation-oriented leaders in securing use of evaluation findings. Based on that experience, I later worked out with Arlen Gullickson the feedback workshop procedure (Gullickson & Stufflebeam, 2001).
8. After expanding The Evaluation Center's work to a range of disciplines, WMU's president and I determined that the Center would function best as a university-level unit—drawing from the university's full range of disciplines. The relocation of the Center was a key factor in my being able to lead the Center and the Colleges of Arts and Sciences, Education, Engineering and Applied Sciences, and Health and Human Services to collaborate in developing WMU's new Interdisciplinary Ph.D. Program in Evaluation (Stufflebeam, (2001-d).
9. Throughout my career I have become increasingly sensitive to evaluation's political nature. In many evaluations different interest groups seek to control, bias, subvert, and/or discredit the evaluation so as to serve their interests. Evaluators must regularly seek, win,

and sustain power over their evaluations to assure their integrity, viability, and credibility. I typically employ a range of measures to help preserve an evaluation's political viability. These include negotiating advance contracts, grounding evaluations in professional standards, engaging stakeholder panels to review evaluation plans and draft reports,⁹ obtaining independent metaevaluations and, when feasible, having a third party fund the metaevaluation (see Finn, Stevens, Stufflebeam, & Walberg (1997).

10. Also, throughout my career I have sought to develop practical evaluation tools. Persons interested in applying the CIPP Model can find at The Evaluation Center's Web site www.wmich.edu/evalctr/checklists a series of checklists on such topics as planning evaluation studies, negotiating evaluation contracts, institutionalizing evaluation, applying professional standards, and designing metaevaluations. The *CIPP Evaluation Model Checklist* is appended to this paper.

The above and many other evaluation experiences at WMU (and Ohio State) have strongly influenced my views of evaluation. As I said at the start of this paper, my ideas about evaluation stem from conducting and thinking about a wide range of evaluation activities and collaborating with evaluation scholars and practitioners, administrators and other staff, and students.

⁹ I have found *review panels* extremely useful. Persons who serve on such panels can make valuable contributions by critiquing evaluation plans and reports from their different perspectives. However, I have an aversion to calling such groups *advisory panels*. Assigning such a label seems to cause persons with no special qualifications in evaluation generally or in the content of the specific study to feel perfectly comfortable and assertive to say how problems seen in the evaluation should be solved. Usually such determinations are best directed to appropriately qualified experts and those with responsibility for carrying out the evaluation.

PART III: PLANNING AND CARRYING THROUGH CIPP EVALUATIONS

This paper's concluding part is keyed to the appended CIPP Evaluation Model Checklist. That checklist is designed to help evaluators and their clients plan, conduct, and assess evaluations based on the requirements of the CIPP Model and the Joint Committee (1994) *Program Evaluation Standards*. While the checklist is self-explanatory and can stand alone in evaluation planning efforts, the following discussion is intended to encourage and support use of the checklist.

The checklist is comprehensive in providing guidance for thoroughly evaluating long-term, ongoing programs. However, users can apply the checklist flexibly and use those parts that fit needs of particular evaluations. Also, the checklist provides guidance for both formative and summative evaluations.

An important feature is the inclusion of checkpoints for both evaluators and clients/stakeholders. For each of the 10 evaluation components, the checklist provides checkpoints on the left for evaluators and corresponding checkpoints on the right for evaluation clients and other users. The checklist thus delineates in some detail what clients and evaluators need to do individually and together to make an evaluation succeed.

Concepts Underlying the Checklist

As seen in this paper's first two parts, the definition of evaluation underlying this checklist is that evaluations should assess and report an entity's merit, worth, probity, and/or significance and also present lessons learned. Moreover, CIPP evaluations and applications of this checklist should meet the Joint Committee (1994) standards of utility, feasibility, propriety, and accuracy. The checklist's contents are configured according to the theme that evaluation's

most important purpose is not to prove, but to improve. Also, as described previously in this paper the recommended evaluation approach is values-based and objective in its orientation.

Contractual Agreements

The checklist's first section identifies essential agreements in negotiating an evaluation contract (or memorandum of agreement). These provide both parties assurances that the evaluation will yield timely, responsive, valid reports and be beyond reproach; necessary cooperation of the client group will be provided; roles of all evaluation participants will be clear; budgetary agreements will be appropriate and clear; and the evaluation agreements will be subject to modification as needed.

CIPP Components

The checklist's next seven sections provide guidance for designing context, input, process, impact, effectiveness, sustainability, and transportability evaluations. Recall that the impact, effectiveness, sustainability, and transportability evaluations are subparts of product evaluation. Experience has shown that such a breakout of product evaluation is important in multiyear evaluations of large scale, long-term programs.

The seven CIPP components may be employed selectively and in different sequences and often simultaneously depending on the needs of particular evaluations. Especially, evaluators should take into account any sound evaluation information the clients/stakeholders already have or can get from other sources. As stressed in Part I of this paper, CIPP evaluations should complement rather than supplant other defensible evaluations of a program or other entity.

Formative Evaluation Reports

Ongoing, formative reporting checkpoints are embedded in each of the CIPP components. These are provided to assist groups to plan, carry out, institutionalize, and/or

disseminate effective services to targeted beneficiaries. Timely communication of relevant, valid evaluation findings to the client and right-to-know audiences is essential in sound evaluations. As needed, findings from the different evaluation components should be drawn together and reported periodically, typically once or twice a year, but more often if needed.

The general process, for each reporting occasion, calls for draft reports to be sent to designated stakeholders about 10 working days prior to a feedback session. At the session the evaluator may use visual aides, e.g., a PowerPoint presentation, to brief the client, staff, and other members of the audience. It is a good idea to provide the client with a copy of the visual aids, so subsequently he or she can brief board members or other stakeholder groups on the most recent evaluation findings. Those present at the feedback session should be invited to raise questions, discuss the findings, and apply them as they choose. At the session's end, the evaluator should summarize the evaluation's planned next steps and future reports; arrange for needed assistance from the client group, especially in data collection; and inquire whether any changes in the data collection and reporting plans and schedule would make future evaluation services more credible and useful. Following the feedback session, the evaluators should finalize the evaluation reports, revise the evaluation plan and schedule as appropriate, and transmit to the client and other designated recipients the finalized reports and any revised evaluation plan and schedule.

Metaevaluation

The checklist's next to last section provides details for both formative and summative metaevaluation. Metaevaluation is to be done throughout the evaluation process. Evaluators should regularly assess their own work against appropriate standards as a means of quality assurance. They should also encourage and cooperate with independent assessments of their

work. Typically, the client or a third party should commission and fund the independent metaevaluation. At the end of the evaluation, evaluators are advised to give their attestation of the extent to which applicable professional standards were met.

The Summative Evaluation Report

The checklist concludes with detailed steps for producing a summative evaluation report. This is a synthesis of all the findings to inform the full range of audiences about what was attempted, done, and accomplished; the bottom-line assessment of the program; and what lessons were learned.

Reporting summative evaluation findings is challenging. A lot of information has to be compiled and communicated effectively. The different audiences likely will have varying degrees of interest and tolerance for long reports. The evaluator should carefully assess the interests and needs of the different audiences and design the final report to help each audience get directly to the information of interest. This checklist recommends that the final report actually be a compilation of three distinct reports.

The first, *program antecedents* report, should inform those not previously acquainted with the program about the sponsoring organization, how and why the program was started, and the environment where it was conducted.

The second, *program implementation* report, should give accurate details of the program to groups that might want to carry out a similar program. Key parts of this report should include descriptions of the program's beneficiaries, goals, procedures, budget, staff, facilities, etc. This report essentially should be objective and descriptive. While it is appropriate to identify important program deficiencies, judgments mainly should be reserved for the *program results* report.

The third, *program results* report, should address questions of interest to all members of the audience. It should summarize the employed evaluation design and procedures. It should then inform all members of the audience about the program's context, input, process, impact, effectiveness, sustainability, and transportability. It should present conclusions on the program's merit, worth, probity, and significance. It should lay out the key lessons learned.

The summative evaluation checkpoint further suggests that, when appropriate, each of the three subreports end with photographs that retell the subreport's account. These can enhance the reader's interest, highlight the most important points, and make the narrative more convincing. A set of photographs (or charts) at the end of each subreport also helps make the overall report seem more approachable than a single, long presentation of narrative. This final checkpoint also suggests interspersing direct quotations from stakeholders to help capture the reader's interest, providing an executive summary for use in policy briefing sessions, and issuing an appendix of evaluation materials to document and establish credibility for the employed evaluation procedures.

SUMMATION

The CIPP Model treats evaluation as an essential concomitant of improvement and accountability within a framework of appropriate values and a quest for clear, unambiguous answers. It responds to the reality that evaluations of innovative, evolving efforts typically cannot employ controlled, randomized experiments or work from published evaluation instruments—both of which yield far too little information anyway. The CIPP Model is configured to enable and guide comprehensive, systematic examination of efforts that occur in

the dynamic, septic conditions of the real world, not the controlled conditions of experimental psychology and split plot crop studies in agriculture.

The model sees evaluation as essential to society's progress and well-being. It contends that societal groups cannot make their programs, services, and products better unless they learn where they are weak and strong. Developers and service providers cannot

- be sure their goals are worthy unless they validate the goals' consistency with sound values and responsiveness to beneficiaries' needs
- plan effectively and invest their time and resources wisely if they don't identify and assess options
- earn continued respect and support if they cannot show that they have responsibly carried out their plans and produced beneficial results
- build on past experiences if they don't preserve, study, and act upon lessons from failed and successful efforts
- convince consumers to buy or support their services and products unless their claims for these services are valid and honestly reported

Institutional personnel cannot meet all of their evaluation needs if they don't both contract for external evaluations and also build and apply capacity to conduct internal evaluations. Evaluators cannot defend their evaluative conclusions unless they key them to sound information and clear, defensible values. Moreover, internal and external evaluators cannot maintain credibility for their evaluations if they do not subject them to metaevaluations against appropriate standards.

The CIPP Model employs multiple methods, is based on a wide range of applications, is keyed to professional standards for evaluations, is supported by an extensive literature, and is

buttressed by practical procedures, including a set of evaluation checklists and particularly the CIPP Evaluation Model Checklist appended to this paper. It cannot be overemphasized, however, that the model is and must be subject to continuing assessment and further development.

References

- Adams, J. A. (1971). *A study of the status, scope, and nature of educational evaluation in Michigan's public K-12 school districts*. Unpublished doctoral dissertation, The Ohio State University, Columbus.
- Alkin, M., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Braybrooke, D., & Lindblom, C. E. (1963). *A strategy of decision*. New York: The Free Press.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.
- Candoli, I. C., Cullen, K., & Stufflebeam, D. L. (1997). *Superintendent performance evaluation: Current practice and directions for improvement*. Boston: Kluwer.
- Cook, D. L., & Stufflebeam, D. L. (1967). Estimating test norms from variable size item and examinee samples. *Educational and Psychological Measurement*, 27, 601-610.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornick, R. C., Philips, D. C., Walker, D. R., & Weiner, S. S. (1981). *Toward reforms of program evaluation*. San Francisco: Jossey-Bass.
- Finn, C. E., Stevens, F. I., Stufflebeam, D. L., & Walberg, H. (1997). In H. L. Miller (Ed.), *The New York City public schools integrated learning systems project: Evaluation and meta-evaluation*. *International Journal of Educational Research*, 27(2), pp. 159-174.
- Gally, J. (1984, April). *The evaluation component*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Granger, A., Grierson, J., Quirino, T. R., & Romano (1965). *Training in planning, monitoring, and evaluation for agricultural research management: Manual 4—Evaluation*. The Hague: International Service for National Agricultural Research.
- Guba, E. G. (1966, October). *A study of Title III Activities; report on evaluation*. National Institute for the Study of Educational Change, Indiana University. (mimeo)
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guba, E. G., & Stufflebeam, D. L. (1968). Evaluation: The process of stimulating, aiding, and abetting insightful action. In R. Ingle & W. Gephart (Eds.), *Problems in the training of educational researchers*. Bloomington, IN: Phi Delta Kappa.
- Guba, E. G., & Stufflebeam, D. L. (1970, June). *Strategies for the institutionalization of the CIPP evaluation model*. An address delivered at the 11th Phi Delta Kappa Symposium on Educational Research, Columbus, OH.
- Gullickson, A., & Stufflebeam, D. (2001, December). *Feedback workshop checklist*. Available: Western Michigan University Evaluation Center Web site: <http://www.wmich.edu/evalctr/checklists/>
- Heck, J., Stufflebeam, D. L., & Hock, M. (1966). *Analysis of educational changes in Ohio public schools*. Columbus: The Ohio State University Press.

- House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation in practice. In G. F. Madaus, D. L. Stufflebeam, & T. Kellaghan (Eds.) (2000), *Evaluation models* (pp. 409-421). Boston: Kluwer.
- House, E. R., Rivers, W., & Stufflebeam, D. L. (1974). An assessment of the Michigan accountability system. *Phi Delta Kappan*, 60(10).
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards*. Thousand Oaks, CA: Corwin.
- Nevo, D. (1974). *Evaluation priorities of students, teachers, and principals*. Unpublished doctoral dissertation, The Ohio State University, Columbus.
- Owens, T. R., & Stufflebeam, D. L. (1964). An experimental comparison of item sampling and examinee sampling for estimating test norms. *Journal of Educational Measurement*, 6(2), 75-83.
- Patton, M. Q. (2000). *Utilization focused evaluation*. In G. M. Madaus, D. L. Stufflebeam, & T. Kellaghan (eds.), *Evaluation models* (pp. 425-438). Boston: Kluwer.
- Reinhard, D. (1972). *Methodology development for input evaluation using advocate and design teams*. Unpublished doctoral dissertation, The Ohio State University, Columbus.
- Scriven, M. (1972). An introduction to metaevaluation. In P. A. Taylor & D. M. Cowley (Eds.), *Readings in curriculum evaluation*. Dubuque, IA: W. C. Brown.
- Shadish, W. R., Newman, D. L., Scheirer, M. A., & Wye, C. (Eds.) (1995). Guiding principles for evaluators. *New Directions for Program Evaluation*, 66.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Boston: Kluwer.
- Smith, M. F. (1989). *Evaluability assessment: A practical approach*. Boston: Kluwer.
- Smith, E. R., & Tyler, R. W. (1942). *Appraising and recording student progress*. New York: Harper.
- Stake, R. (1983). Program evaluation, particularly responsive evaluation. In G. Madaus, M. Scriven, and D. Stufflebeam (Eds.), *Evaluation models*. Boston: Kluwer-Nijhoff.
- Stake, R. (2001). *Checklist for negotiating an agreement to evaluate an educational programme*. Available: Western Michigan University Evaluation Center Web site: www.wmich.edu/evalctr/checklists.
- Stufflebeam, D. L. (1966-a, January). *Evaluation under Title I of the Elementary and Secondary Education Act of 1967*. Address delivered at evaluation conference sponsored by the Michigan State Department of Education.
- Stufflebeam, D. L. (1966-b). A depth study of the evaluation requirement. *Theory Into Practice*, 5(3), 121-133.
- Stufflebeam, D. L. (1967-a). The use and abuse of evaluation in Title III. *Theory Into Practice*, 6, 126-133.
- Stufflebeam, D. L. (1967-b). *Applying PERT to a test development project*. Paper presented at the annual meeting of the American Educational Research Association.

- Stufflebeam, D. L. (1968, January). *Evaluation as enlightenment for decision-making*. Paper presented at the Association for Supervision and Curriculum Development Conference on Assessment Theory, Sarasota.
- Stufflebeam, D. L. (1969). Evaluation as enlightenment for decision-making. In A. Walcott (Ed.), *Improving educational assessment and an inventory of measures of affective behavior*. Washington, DC: Association for Supervision and Curriculum Development.
- Stufflebeam, D. L. (1974). *Meta-evaluation*. Occasional Paper Series, Paper No. 3, Kalamazoo: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. (1982). Institutional self-evaluation. *International encyclopedia of educational research*. Oxford, England: Pergamon Press.
- Stufflebeam, D. L. (1995). Evaluation of superintendent performance: Toward a general model. In A. McConney (Eds.), *Toward a unified model of educational personnel evaluation*. Kalamazoo: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. (1997-a). *Strategies for institutionalizing evaluation: revisited*. Occasional Paper Series #18. Kalamazoo: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. (1997-b). A standards-based perspective on evaluation. In R. E. Stake & L. Mabry (Eds.), *Advances in program evaluation, Vol. 3: Evaluation and the postmodern dilemma* (pp. 61-88). Greenwich, CT: JAI Press Inc.
- Stufflebeam, D. L. (1999-a). Contracting for evaluations. *Evaluation: News & Comment* 8(1), 16-17.
- Stufflebeam, D. L. (1999-b). Using professional standards to legally and ethically release evaluation findings. *Studies in Educational Evaluation*, 25(4), 325-334.
- Stufflebeam, D. L. (2000). Lessons in contracting for evaluations. *American Journal of Evaluation*, 21(3), 293-314.
- Stufflebeam, D. L. (2001-a). *Evaluation contracts checklist*. Available: Western Michigan University Web site: www.wmich.edu/evalctr/checklists.
- Stufflebeam, D. L. (2001-b). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L. (2001-c, Spring). Evaluation models. *New Directions for Evaluation*, 89, 7-99.
- Stufflebeam, D. L. (2001-d). Interdisciplinary Ph.D. programming in evaluation. *American Journal of Evaluation*, 22(3), 445-455.
- Stufflebeam, D. L. (2002-a). *CIPP evaluation model checklist*. Available: Western Michigan University Evaluation Center Web site: www.wmich.edu/evalctr/checklists.
- Stufflebeam, D. L. (2002-b). *University-based R & D centers checklist*. Available: Western Michigan University Evaluation Center Web site: www.wmich.edu/evalctr/checklists.
- Stufflebeam, D. L. (2003-a). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *The international handbook of educational evaluation* (Chapter 3). Boston: Kluwer.
- Stufflebeam, D. L. (2003-b). Institutionalizing evaluation in schools. In T. Kellaghan & D. L. Stufflebeam (Eds.), *The international handbook of educational evaluation* (Chapter 34). Boston: Kluwer.
- Stufflebeam, D. L., Candoli, C., & Nicholls, C. (1995). *A portfolio for evaluation of school superintendents*. Kalamazoo: Center for Research on Educational Accountability and Teacher Evaluation, The Evaluation Center, Western Michigan University.

- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.
- Stufflebeam, D. L., Gullickson, A. R., & Wingate, L. A. (2002). *The spirit of Consuelo: An evaluation of Ke Aka Ho'ona*. Kalamazoo: Western Michigan University Evaluation Center.
- Stufflebeam, D. L., Jaeger, R. M., & Scriven, M. (1992, April 21). *A retrospective analysis of a summative evaluation of NAGB's pilot project to set achievement levels on the national assessment of educational progress*. Chair and presenter at annual meeting of the American Educational Research Association, San Francisco.
- Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer.
- Stufflebeam, D. L., & Millman, J. (1995, December). A proposed model for superintendent evaluation. *Journal of Personnel Evaluation in Education*, 9(4), 383-410.
- Stufflebeam, D. L., & Nevo, D. (1976, Winter). The availability and importance of evaluation information within the school. *Studies in Educational Evaluation*, 2, 203-9.
- Stufflebeam, D. L., & Webster, W. J. (1988). Evaluation as an administrative function. In N. Boyan (Ed.), *Handbook of research on educational administration* (pp. 569-601). White Plains, NY: Longman.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 36, 492-501.
- U. S. General Accounting Office. (2003). *Government auditing standards (The yellow book)*. Washington DC: Author.
- U. S. Office of Education. (1966). *Report of the first year of Title I of the Elementary and Secondary Education Act*. Washington, DC: General Accounting Office.
- Webster, W. J. (1975, March). *The organization and functions of research and evaluation in large urban school districts*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

APPENDIX

CIPP Evaluation Model Checklist

CIPP EVALUATION MODEL CHECKLIST

A tool for applying the *CIPP Model* to assess long-term enterprises

Intended for use by evaluators and evaluation clients/stakeholders

Daniel L. Stufflebeam

August 2003

The CIPP Evaluation Model is a comprehensive framework for guiding evaluations of programs, projects, personnel, products, institutions, and systems. This checklist, patterned after the CIPP Model, is focused on program evaluations, particularly those aimed at effecting long-term, sustainable improvements.

The checklist especially reflects the eight-year evaluation (1994-2002), conducted by the Western Michigan University Evaluation Center, of Consuelo Foundation's values-based, self-help housing and community development program—named Ke Aka Ho'ona—for low income families in Hawaii. Also, It is generally consistent with a wide range of program evaluations conducted by The Evaluation Center in such areas as science and mathematics education, rural education, educational research and development, achievement testing, state systems of educational accountability, school improvement, professional development schools, transition to work, training and personnel development, welfare reform, nonprofit organization services, community development, community-based youth programs, community foundations, and technology.

Corresponding to the letters in the acronym CIPP, this model's core parts are context, input, process, and product evaluation. In general, these four parts of an evaluation respectively ask, What needs to be done? How should it be done? Is it being done? Did it succeed?

In this checklist, the “Did it succeed?” or product evaluation part is divided into impact, effectiveness, sustainability, and transportability evaluations. Respectively, these four product evaluation subparts ask, Were the right beneficiaries reached? Were their needs met? Were the gains for the beneficiaries sustained? Did the processes that produced the gains prove transportable and adaptable for effective use in other settings?

This checklist represents a recent update of the CIPP Model. The model's first installment—actually before all 4 CIPP parts were introduced—was published more than 35 years ago (Stufflebeam, 1966) and stressed the need for process as well as product evaluations. The second installment—published a year later (Stufflebeam, 1967)—included context, input, process, and product evaluations and emphasized that goal-setting should be guided by context evaluation, including a needs assessment, and that program planning should be guided by input evaluation, including assessments of alternative program strategies. The third installment (Stufflebeam, D. L., Foley, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M., 1971) set the 4 types of evaluation within a systems, improvement-oriented framework. The model's fourth installment (Stufflebeam, 1972) showed how the model could and should be used for summative as well as formative evaluation. The model's fifth installment—illustrated by this checklist—breaks out product evaluation into the above-noted four subparts in order to help assure and

assess a program's long-term viability. (See Stufflebeam, in press-a and -b.)

This checklist is designed to help evaluators evaluate programs with relatively long-term goals. The checklist's first main function is to provide timely evaluation reports that assist groups to plan, carry out, institutionalize, and/or disseminate effective services to targeted beneficiaries. The checklist's other main function is to review and assess a program's history and to issue a summative evaluation report on its merit, worth, and significance and the lessons learned.

This checklist has 10 components. The first—contractual agreements to guide the evaluation—is followed by the context, input, process, impact, effectiveness, sustainability, and transportability evaluation components. The last 2 are metaevaluation and the final synthesis report. Contracting for the evaluation is done at the evaluation's outset, then updated as needed. The 7 CIPP components may be employed selectively and in different sequences and often simultaneously depending on the needs of particular evaluations. Especially, evaluators should take into account any sound evaluation information the clients/stakeholders already have or can get from other sources. CIPP evaluations should complement rather than supplant other defensible evaluations of an entity. Metaevaluation (evaluation of an evaluation) is to be done throughout the evaluation process; evaluators also should encourage and cooperate with independent assessments of their work. At the end of the evaluation, evaluators are advised to give their attestation of the extent to which applicable professional standards were met. This checklist's final component provides concrete advice for compiling the final summative evaluation report, especially by drawing together the formative evaluation reports that were issued throughout the evaluation.

The concept of evaluation underlying the CIPP Model and this checklist is that evaluations should assess and report an entity's merit, worth, probity, and significance and also present lessons learned. Moreover, CIPP evaluations and applications of this checklist should meet the Joint Committee (1994) standards of utility, feasibility, propriety, and accuracy. The model's main theme is that evaluation's most important purpose is not to prove, but to improve.

Timely communication of relevant evaluation findings to the client and right-to-know audiences is another key theme of this checklist. As needed, findings from the different evaluation components should be drawn together and reported periodically, typically once or twice a year. The general process, for each reporting occasion, calls for draft reports to be sent to designated stakeholders about 10 days prior to a feedback workshop.² At the workshop the evaluators should use visual aids, e.g., a PowerPoint presentation to brief the client, staff, and other members of the audience. (It is often functional to provide the clients with a copy of the visual aids, so subsequently they can brief members of their boards or other stakeholder groups on the most recent evaluation findings.) Those present at the feedback workshop should be invited to raise questions, discuss the findings, and apply them as they choose. At the workshop's end, the evaluators should summarize the evaluation's planned next steps and future reports; arrange for needed assistance from the client group, especially in data collection; and inquire whether any changes in the data collection and reporting plans and schedule would make future evaluation services more credible and useful. Following the feedback workshop, the evaluators should finalize the evaluation reports, revise the evaluation plan and schedule as appropriate, and transmit to the client and other designated recipients the finalized reports and any revised evaluation plans and schedule.

Beyond guiding the evaluator's work, the checklist gives advice for evaluation users. For each of the 10 evaluation components, the checklist provides checkpoints on the left for evaluators and checkpoints on the right for evaluation clients and other users.

For more information about the CIPP Model, please consult the references and related checklists listed at the end of this checklist.

1. CONTRACTUAL AGREEMENTS

CIPP evaluations should be grounded in explicit advance agreements with the client, and these should be updated as needed throughout the evaluation. (See Daniel Stuffelbeam's *Evaluation Contracts Checklist* at www.wmich.edu/evalctr/checklists)

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Contracting</i>
<input type="checkbox"/> Develop a clear understanding of the evaluation job to be done.	<input type="checkbox"/> Clarify with the evaluator what is to be evaluated, for what purpose, according to what criteria, and for what audiences.
<input type="checkbox"/> Secure agreements needed to assure that the right information can be obtained.	<input type="checkbox"/> Clarify with the evaluator what information is essential to the evaluation and how the client group will facilitate its collection.
<input type="checkbox"/> Clarify for the client, in general, what quantitative and qualitative analyses will be needed to make a full assessment of the program.	<input type="checkbox"/> Reach agreements with the evaluator on what analyses will be most important in addressing the client group's questions.
<input type="checkbox"/> Clarify the nature, general contents, and approximate required timing of the final summative evaluation report.	<input type="checkbox"/> Assure that the planned final report will meet the needs of the evaluation's different audiences.
<input type="checkbox"/> Clarify the nature, general contents, and timing of interim, formative evaluation reports and reporting sessions.	<input type="checkbox"/> Assure that the evaluation's reporting plan and schedule are functionally responsive to the needs of the program.
<input type="checkbox"/> Reach agreements to protect the integrity of the reporting process.	<input type="checkbox"/> Assure that the reporting process will be legally, politically, and ethically viable.
<input type="checkbox"/> Clarify the needed channels for communication and assistance from the client and other stakeholders.	<input type="checkbox"/> Assure that the evaluation plan is consistent with the organization's protocol.
<input type="checkbox"/> Secure agreements on the evaluation's time line and who will carry out the evaluation responsibilities.	<input type="checkbox"/> Clarify for all concerned parties the evaluation roles and responsibilities of the client group.
<input type="checkbox"/> Secure agreements on the evaluation budget and payment amounts and dates.	<input type="checkbox"/> Assure that budgetary agreements are clear and functionally appropriate for the evaluation's success.
<input type="checkbox"/> Clearly define provisions for reviewing, controlling, amending, and/or canceling the evaluation.	<input type="checkbox"/> Assure that the evaluation will be periodically reviewed and, as needed and appropriate, subject to modification and termination.

2. CONTEXT EVALUATION

Context evaluation assesses needs, assets, and problems within a defined environment.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Program Aims</i>
<input type="checkbox"/> Compile and assess background information, especially on the intended beneficiaries' needs and assets.	<input type="checkbox"/> Use the context evaluation findings in selecting and/or clarifying the intended beneficiaries.
<input type="checkbox"/> Interview program leaders to review and discuss their perspectives on beneficiaries' needs and to identify any problems (political or otherwise) the program will need to solve.	<input type="checkbox"/> Use the context evaluation findings in reviewing and revising, as appropriate, the program's goals to assure they properly target assessed needs.
<input type="checkbox"/> Interview other stakeholders to gain further insight into the needs and assets of intended beneficiaries and potential problems for the program.	<input type="checkbox"/> Use the context evaluation findings in assuring that the program is taking advantage of pertinent community and other assets.
<input type="checkbox"/> Assess program goals in light of beneficiaries' assessed needs and potentially useful assets.	<input type="checkbox"/> Use the context evaluation findings—throughout and at the program's end—to help assess the program's effectiveness and significance in meeting beneficiaries' assessed needs.
<input type="checkbox"/> Engage an evaluator ³ to monitor and record data on the program's environment, including related programs, area resources, area needs and problems, and political dynamics.	
<input type="checkbox"/> Request that program staff regularly make available to the evaluation team information they collect on the program's beneficiaries and environment.	
<input type="checkbox"/> Annually, or as appropriate, prepare and deliver to the client and agreed-upon stakeholders a draft context evaluation report providing an update on program-related needs, assets, and problems, along with an assessment of the program's goals and priorities.	
<input type="checkbox"/> Discuss context evaluation findings in feedback workshops presented about annually to the client and designated audiences.	
<input type="checkbox"/> Finalize context evaluation reports and associated visual aids and provide them to the client and agreed-upon stakeholders. ⁴	

3. INPUT EVALUATION

Input evaluation assesses competing strategies and the work plans and budgets of the selected approach.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Program Planning</i>
<input type="checkbox"/> Identify and investigate existing programs that could serve as a model for the contemplated program.	<input type="checkbox"/> Use the input evaluation findings to devise a program strategy that is scientifically, economically, socially, politically, and technologically defensible.
<input type="checkbox"/> Assess the program's proposed strategy for responsiveness to assessed needs and feasibility.	<input type="checkbox"/> Use the input evaluation findings to assure that the program's strategy is feasible for meeting the assessed needs of the targeted beneficiaries.
<input type="checkbox"/> Assess the program's budget for its sufficiency to fund the needed work.	<input type="checkbox"/> Use the input evaluation findings to support funding requests for the planned enterprise.
<input type="checkbox"/> Assess the program's strategy against pertinent research and development literature.	<input type="checkbox"/> Use the input evaluation findings to train staff to carry out the program.
<input type="checkbox"/> Assess the merit of the program's strategy compared with alternative strategies found in similar programs.	<input type="checkbox"/> Use the input evaluation findings for accountability purposes in reporting the rationale for the selected program strategy and the defensibility of the operational plan.
<input type="checkbox"/> Assess the program's work plan and schedule for sufficiency, feasibility, and political viability.	
<input type="checkbox"/> Compile a draft input evaluation report and send it to the client and agreed-upon stakeholders.	
<input type="checkbox"/> Discuss input evaluation findings in a feedback workshop.	
<input type="checkbox"/> Finalize the input evaluation report and associated visual aids and provide them to the client and agreed-upon stakeholders.	

4. PROCESS EVALUATION

Process evaluations monitor, document, and assess program activities.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Managing and Documenting</i>
<input type="checkbox"/> Engage an evaluation team member to monitor, observe, maintain a photographic record of, and provide periodic progress reports on program implementation.	<input type="checkbox"/> Use the process evaluation findings to control and strengthen staff activities.
<input type="checkbox"/> In collaboration with the program's staff, maintain a record of program events, problems, costs, and allocations.	<input type="checkbox"/> Use the process evaluation findings to strengthen the program design.
<input type="checkbox"/> Periodically interview beneficiaries, program leaders, and staff to obtain their assessments of the program's progress.	<input type="checkbox"/> Use the process evaluation findings to maintain a record of the program's progress.
<input type="checkbox"/> Maintain an up-to-date profile of the program.	<input type="checkbox"/> Use the process evaluation findings to help maintain a record of the program's costs.
<input type="checkbox"/> Periodically draft written reports on process evaluation findings and provide the draft reports to the client and agreed-upon stakeholders.	<input type="checkbox"/> Use the process evaluation findings to report on the program's progress to the program's financial sponsor, policy board, community members, other developers, etc.
<input type="checkbox"/> Present and discuss process evaluation findings in feedback workshops.	
<input type="checkbox"/> Finalize each process evaluation report (possibly incorporated into a larger report) and associated visual aids and provide them to the client and agreed-upon stakeholders.	

5. IMPACT EVALUATION

Impact evaluation assesses a program's reach to the target audience.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Controlling Who Gets Served</i>
<input type="checkbox"/> Engage the program's staff and consultants and/or an evaluation team member to maintain a directory of persons and groups served, make notations on their needs, and record program services they received.	<input type="checkbox"/> Use the impact evaluation findings to assure that the program is reaching intended beneficiaries.
<input type="checkbox"/> Assess and make a judgment of the extent to which the served individuals and groups are consistent with the program's intended beneficiaries.	<input type="checkbox"/> Use the impact evaluation findings to assess whether the program is reaching or did reach inappropriate beneficiaries.
<input type="checkbox"/> Periodically interview area stakeholders, such as community leaders, employers, school and social programs personnel, clergy, police, judges, and homeowners, to learn their perspectives on how the program is influencing the community.	<input type="checkbox"/> Use the impact evaluation findings to judge the extent to which the program is serving or did serve the right beneficiaries.
<input type="checkbox"/> Include the obtained information and the evaluator's judgments in a periodically updated program profile.	<input type="checkbox"/> Use the impact evaluation findings to judge the extent to which the program addressed or is addressing important community needs.
<input type="checkbox"/> Determine the extent to which the program reached an appropriate group of beneficiaries.	<input type="checkbox"/> Use the impact evaluation findings for accountability purposes regarding the program's success in reaching the intended beneficiaries.
<input type="checkbox"/> Assess the extent to which the program inappropriately provided services to a nontargeted group.	
<input type="checkbox"/> Draft an impact evaluation report (possibly incorporated into a larger report) and provide it to the client and agreed-upon stakeholders.	
<input type="checkbox"/> Discuss impact evaluation findings in a feedback workshop.	
<input type="checkbox"/> Finalize the impact evaluation report and associated visual aids and provide them to the client and agreed-upon stakeholders.	

6. EFFECTIVENESS EVALUATION

Effectiveness evaluation assesses the quality and significance of outcomes.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Assessing/Reporting Outcomes</i>
<input type="checkbox"/> Interview key stakeholders, such as community leaders, beneficiaries, program leaders and staff, and other interested parties, to determine their assessments of the program's positive and negative outcomes.	<input type="checkbox"/> Use effectiveness evaluation findings to gauge the program's positive and negative effects on beneficiaries.
	<input type="checkbox"/> Use the effectiveness evaluation findings to gauge the program's positive and negative effects on the community/pertinent environment.
<input type="checkbox"/> Conduct in-depth case studies of selected beneficiaries.	<input type="checkbox"/> Use the effectiveness evaluation findings to sort out and judge important side effects.
<input type="checkbox"/> Engage an evaluation team member and program staff to supply documentation needed to identify and confirm the range, depth, quality, and significance of the program's effects on beneficiaries.	<input type="checkbox"/> Use the effectiveness evaluation findings to examine whether program plans and activities need to be changed.
<input type="checkbox"/> Engage an evaluation team member to compile and assess information on the program's effects on the community.	<input type="checkbox"/> Use the effectiveness evaluation findings to prepare and issue program accountability reports.
<input type="checkbox"/> Engage a goal-free evaluator ⁵ to ascertain what the program actually did and to identify its full range of effects—positive and negative, intended and unintended.	<input type="checkbox"/> Use the effectiveness evaluation findings to make a bottom-line assessment of the program's success.
<input type="checkbox"/> Obtain information on the nature, cost, and success of similar programs conducted elsewhere and judge the subject program's effectiveness in contrast to the identified "critical competitors."	<input type="checkbox"/> Use needs assessment data (from the context evaluation findings), effectiveness evaluation findings, and contrasts with similar programs elsewhere to make a bottom-line assessment of the program's significance.
<input type="checkbox"/> Compile effectiveness evaluation findings in a draft report (that may be incorporated in a larger report) and present it to the client and agreed-upon stakeholders.	
<input type="checkbox"/> Discuss effectiveness evaluation findings in a feedback workshop.	
<input type="checkbox"/> Finalize the effectiveness evaluation report and present it to the client and agreed-upon stakeholders.	
<input type="checkbox"/> Incorporate the effectiveness evaluation findings in an updated program profile and ultimately in the final evaluation report.	

7. SUSTAINABILITY EVALUATION

Sustainability evaluation assesses the extent to which a program's contributions are successfully institutionalized and continued over time.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities: Continuing Successful Practices</i>
<input type="checkbox"/> Interview program <i>leaders</i> and staff to identify their judgments about what program successes should be sustained.	<input type="checkbox"/> Use the sustainability evaluation findings to determine whether staff and beneficiaries favor program continuation.
<input type="checkbox"/> Interview program <i>beneficiaries</i> to identify their judgments about what program successes should be sustained.	<input type="checkbox"/> Use the sustainability findings to assess whether there is a continuing need/demand and compelling case for sustaining the program's services.
<input type="checkbox"/> Review the evaluation's data on program effectiveness, program costs, and beneficiary needs to judge what program successes should and can be sustained.	<input type="checkbox"/> Use the sustainability findings as warranted to set goals and plan for continuation activities.
<input type="checkbox"/> Interview <i>beneficiaries</i> to identify their understanding and assessment of the program's provisions for continuation.	<input type="checkbox"/> Use the sustainability findings as warranted to help determine how best to assign authority and responsibility for program continuation.
<input type="checkbox"/> Obtain and examine plans, budgets, staff assignments, and other relevant information to gauge the likelihood that the program will be sustained.	<input type="checkbox"/> Use the sustainability findings as warranted to help plan and budget continuation activities.
<input type="checkbox"/> Periodically revisit the program to assess the extent to which its successes are being sustained.	
<input type="checkbox"/> Compile and report sustainability findings in the evaluation's progress and final reports.	
<input type="checkbox"/> In a feedback workshop, discuss sustainability findings plus the possible need for a follow-up study to assess long-term results.	
<input type="checkbox"/> Finalize the sustainability evaluation report and present it to the client and agreed-upon stakeholders.	

8. TRANSPORTABILITY EVALUATION

Transportability evaluation assesses the extent to which a program has (or could be) successfully adapted and applied elsewhere.

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities—Dissemination</i>
<input type="checkbox"/> Engage the program staff in identifying actual or potential adopters of the program by keeping a log of inquiries, visitors, and adaptations of the program.	<input type="checkbox"/> Use the transportability evaluation findings to assess the need for disseminating information on the program.
<input type="checkbox"/> Survey a representative sample of potential adopters. Ask them to (1) review a description of the program and a summary of evaluation findings; (2) judge the program's relevance to their situation; (3) judge the program's quality, significance, and replicability; and (4) report whether they are using or plan to adopt all or parts of the program.	<input type="checkbox"/> Use the transportability evaluation findings to help determine audiences for information on the program.
	<input type="checkbox"/> Use the transportability evaluation findings to help determine what information about the program should be disseminated.
	<input type="checkbox"/> Use the transportability evaluation findings to gauge how well the program worked elsewhere.
<input type="checkbox"/> Visit and assess adaptations of the program.	
<input type="checkbox"/> Compile and report transportability evaluation findings in draft reports.	
<input type="checkbox"/> Discuss transportability evaluation findings in a feedback workshop.	
<input type="checkbox"/> Finalize the transportability evaluation report and associated visual aids and present them to the client and agreed-upon stakeholders.	

9. META-EVALUATION⁶

Metaevaluation is an assessment of an evaluation's adherence to pertinent standards of sound evaluation (See Stufflebeam, Daniel. *Program Evaluations Metaevaluation Checklist*. www.wmich.edu/evalctr/checklists)

<i>Evaluator Activities</i>	<i>Client/Stakeholder Activities–Judgment of the Evaluation</i>
<input type="checkbox"/> Reach agreement with the client that the evaluation will be guided and assessed against the Joint Committee Program Evaluation Standards of utility, feasibility, propriety, and accuracy and/or some other mutually agreeable set of evaluation standards or guiding principles.	<input type="checkbox"/> Review the Joint Committee Program Evaluation Standards and reach an agreement with the evaluators that these standards and/or other standards and/or guiding principles will be used to guide and judge the evaluation work.
<input type="checkbox"/> Encourage and support the client to obtain an independent assessment of the evaluation plan, process, and/or reports.	<input type="checkbox"/> Consider contracting for an independent assessment of the evaluation.
<input type="checkbox"/> Document the evaluation process and findings, so that the evaluation can be rigorously studied and evaluated.	<input type="checkbox"/> Keep a file of information pertinent to judging the evaluation against the agreed-upon evaluation standards and/or guiding principles.
<input type="checkbox"/> Steadfastly apply the Joint Committee Standards and/or other set of agreed-upon standards or guiding principles to help assure that the evaluation will be sound and fully accountable.	<input type="checkbox"/> Supply information and otherwise assist as appropriate all legitimate efforts to evaluate the evaluation.
<input type="checkbox"/> Periodically use the metaevaluation findings to strengthen the evaluation as appropriate.	<input type="checkbox"/> Raise questions about and take appropriate steps to assure that the evaluation adheres to the agreed-upon standards and/or other standards/guiding principles.
<input type="checkbox"/> Assess and provide written commentary on the extent to which the evaluation ultimately met each agreed-upon standard and/or guiding principle, and include the results in the final evaluation report's technical appendix.	<input type="checkbox"/> Take into account metaevaluation results in deciding how best to apply the evaluation findings.
	<input type="checkbox"/> Consider appending a statement to the final evaluation report reacting to the evaluation, to the evaluators' attestation of the extent to which standards and/or guiding principles were met, to the results of any independent metaevaluation, and also documenting significant uses of the evaluation findings.

10. THE FINAL SYNTHESIS REPORT

Final synthesis reports pull together evaluation findings to inform the full range of audiences about what was attempted, done, and accomplished; what lessons were learned; and the bottom-line assessment of the program.

Evaluator Activities	Client/Stakeholder Activities: Summing Up
<input type="checkbox"/> Organize the report to meet the differential needs of different audiences, e.g., provide three reports in one, including program antecedents, program implementation, and program results.	<input type="checkbox"/> Help assure that the planned report contents will appeal to and be usable by the full range of audiences.
<input type="checkbox"/> Continuing the example, in the <i>program antecedents</i> report include discrete sections on the organization that sponsored the program, the origin of the program being evaluated, and the program's environment.	<input type="checkbox"/> Help assure that the historical account presented in the <i>program antecedents</i> report is accurate, sufficiently brief, and of interest and use to at least some of the audiences for the overall report.
<input type="checkbox"/> In the <i>program implementation</i> report include sections that give detailed accounts of how the main program components were planned, funded, staffed, and carried out such that groups interested in replicating the program could see how they might conduct the various program activities. These sections should be mainly descriptive and evaluative only to the extent of presenting pertinent cautions.	<input type="checkbox"/> Help assure that the account of program implementation is accurate and sufficiently detailed to help others understand and possibly apply the program's procedures (taking into account pertinent cautions).
	<input type="checkbox"/> Use the <i>program results</i> report to take stock of what was accomplished, what failures and shortfalls occurred, how the effort compares with similar programs elsewhere, and what lessons should be heeded in future programs.
<input type="checkbox"/> In the <i>program results</i> report include sections on the evaluation design, the evaluation findings (divided into <i>context, input, process, impact, effectiveness, sustainability, and transportability</i>), and the evaluation conclusions (divided into <i>strengths, weaknesses, lessons learned, and bottom-line assessment of the program's merit, worth, and significance</i>). Contrast the program's contributions with what was intended, what the beneficiaries needed, what the program cost, and how it compares with similar programs elsewhere.	<input type="checkbox"/> Use the full report as a means of preserving institutional memory of the program and informing interested parties about the enterprise.
<input type="checkbox"/> At the end of each of the three reports, include photographs and graphic representations that help retell the report's particular accounts.	
<input type="checkbox"/> Supplement the main report contents with pithy, pertinent quotations, throughout; a prologue recounting how the evaluation was initiated; an epilogue identifying needed further program and evaluation efforts; an executive summary; acknowledgements; information about the evaluators; and technical appendices containing such items as interview protocols and questionnaires.	

BIBLIOGRAPHY

- Controller General of the United States. (2002, January). *Government auditing standards* (2002 revision, exposure draft–GAO-02-340G). Washington, DC: U.S. General Accounting Office.
- Guba, E. G., & Stufflebeam, D. L. (1968). Evaluation: The process of stimulating, aiding, and abetting insightful action. In R. Ingle & W. Gephart (Eds.), *Problems in the training of educational researchers*. Bloomington, IN: Phi Delta Kappa.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.
- Shadish, W. R., Newman, D. L., Scheirer, M. A., & Wye, C. (1995). Guiding principles for evaluators. *New Directions for Program Evaluation*, 66.
- Stufflebeam, D. L. (1966). A depth study of the evaluation requirement. *Theory Into Practice*, 5(3), 121-133.
- Stufflebeam, D. L. (1967, June). The use and abuse of evaluation in Title III. *Theory Into Practice* 6, 126-133.
- Stufflebeam, D. L. (1969). Evaluation as enlightenment for decision-making. In H. B. Walcott (Ed.), *Improving educational assessment and an inventory of measures of affective behavior* (pp. 41-73). Washington, DC: Association for Supervision and Curriculum Development and National Education Association.
- Stufflebeam, D. L. (1972). The relevance of the CIPP evaluation model for educational accountability. *SRIS Quarterly*, 5(1).
- Stufflebeam, D. L. (1973). Evaluation as enlightenment for decision-making. In B. R. Worthen & J. R. Sanders (Eds.), *Educational evaluation: Theory and practice*. Worthington, OH: Charles A. Jones Publishing Company.
- Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models* (Chapter 7, pp. 117-141). Boston: Kluwer-Nijhoff.
- Stufflebeam, D. L. (1985). Stufflebeam's improvement-oriented evaluation. In D. L. Stufflebeam & A. J. Shinkfield (Eds.), *Systematic evaluation* (Chapter 6, pp. 151-207). Boston: Kluwer-Nijhoff.
- Stufflebeam, D. L. (1997). *Strategies for institutionalizing evaluation: revisited*. Occasional Paper Series #18. Kalamazoo, MI: Western Michigan University Evaluation Center. (May be purchased from The Evaluation Center for \$5. To place an order, contact Eval-Center@wmich.edu)
- Stufflebeam, D.L. (2000). The CIPP model for evaluation. In D.L. Stufflebeam, G. F. Madaus, & T. Kellaghan, (Eds.), *Evaluation models* (2nd ed.). (Chapter 16). Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.

Stufflebeam, D. L. (in press-a). The CIPP model for evaluation. In D. L. Stufflebeam, & T. Kellaghan, (Eds.), *The international handbook of educational evaluation* (Chapter 2). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L. (in press-b). Institutionalizing evaluation in schools. In D. L. Stufflebeam, & T. Kellaghan, (Eds.), *The international handbook of educational evaluation* (Chapter 34). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. (1971). *Educational evaluation and decision making* (Chapters 3, 7, & 8). Itasca, IL: F. E. Peacock.

Stufflebeam, D. L., & Webster, W. J. (1988). Evaluation as an administrative function. In N. Boyan (Ed.), *Handbook of research on educational administration* (pp. 569-601). White Plains, NY: Longman.

RELATED CHECKLISTS

(available at www.wmich.edu/evalctr/checklists)

Checklist for Negotiating an Agreement to Evaluate an Educational Program by Robert Stake

Checklist for Developing and Evaluating Evaluation Budgets by Jerry Horn

Evaluation Contracts Checklist by Daniel Stufflebeam

Evaluation Plans and Operations Checklist by Daniel Stufflebeam

Evaluation Values and Criteria Checklist by Daniel Stufflebeam

Feedback Workshop Checklist by Arlen Gullickson & Daniel Stufflebeam

Guiding Principles Checklist by Daniel Stufflebeam

Program Evaluations Metaevaluation Checklist (Based on *The Program Evaluation Standards*) by Daniel Stufflebeam

Notes

2. The feedback workshops referenced throughout the checklist are a systematic approach by which evaluators present, discuss, and examine findings with client groups. A checklist for planning feedback workshops can be found at www.wmich.edu/evalctr/checklists/.
3. Applications of the CIPP Model have typically included evaluation team members who spend much time at the program site systematically observing and recording pertinent information. Called Traveling Observers when program sites are dispersed or Resident Observers when program activities are all at one location, these evaluators help design and subsequently work from a specially constructed Traveling Observer's Handbook containing prescribed evaluation questions, procedures, forms, and reporting formats. Such handbooks are tailored to the needs of the particular evaluation. While the observers focus heavily on context and process evaluations, they may also collect and report information on program plans, costs, impacts, effectiveness, sustainability, and transportability.
4. Whereas each of the seven evaluation components includes a reporting function, findings from the different components are not necessarily presented in separate reports. Depending on the circumstances of a particular reporting occasion, availability of information from different evaluation components, and the needs and preferences of the audience, information across evaluation components may be combined in one or more composite reports. Especially, process, impact, and effectiveness information are often combined in a single report. The main point is to design and deliver evaluation findings so that the audience's needs are served effectively and efficiently.
5. A goal-free evaluator is a contracted evaluator who, by agreement, is prevented from learning a program's goals and is charged to assess what the program is actually doing and achieving, irrespective of its aims. This technique is powerful for identifying side effects, or unintended outcomes, both positive and negative, also for describing what the program is actually doing, irrespective of its stated procedures.
6. See the RELATED CHECKLISTS section to identify a number of checklists designed to guide metaevaluations.